NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: T-385
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): READINESS EVALUATION USING MULTIVARIATE DATA REDUCTION
5. TYPE OF REPORT & PERIOD COVERED: SCIENTIFIC
6. PERFORMING ORG. REPORT NUMBER: T-385
7. AUTHOR(s): ZEEV BARZILY, W. H. MARLOW, S. ZACKS
8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0729
9. PERFORMING ORGANIZATION NAME AND ADDRESS: THE GEORGE WASHINGTON UNIVERSITY, PROGRAM IN LOGISTICS, WASHINGTON, D. C. 20037
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: OFFICE OF NAVAL RESEARCH, CODE 434, ARLINGTON, VA 22217
12. REPORT DATE: 1 November 1978
13. NUMBER OF PAGES: 60
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): NONE

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MULTIVARIATE ANALYSIS
READINESS
OPERATIONAL READINESS
MATERIAL READINESS
LOGISTICS

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several statistical methods--principal component analysis, orthogonal factor analysis, classification, and clustering techniques--are tailored and combined into a system designed to digest high-dimensional vectors of data on operational readiness of Navy ships. Such data consist of large numbers of scores for individual ships assigned by experts. The purpose of the data reduction system is to provide a robust method of representing the data by a small number of scores that are meaningfully related

Cont'd

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT CONT'D

to the original scores and that allow classification and clustering of the ships into homogeneous groups on relevant readiness scales. Simulated data drawn from mixtures of specified multivariate normal populations have been used to test the ability of the system to recover individual populations and to detect trends over time.

READINESS EVALUATIONS USING
MULTIVARIATE DATA REDUCTION

by

Zeev Barzily
W. H. Marlow
S. Zacks

Serial T-385
1 November 1978

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

Contract N00014-75-C-0729
Project NR 347 020
Office of Naval Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

Abstract
of
Serial T-385
1 November 1978

READINESS EVALUATIONS USING MULTIVARIATE DATA REDUCTION

by

Zeev Barzily
W. H. Marlow
S. Zacks

Several statistical methods--principal component analysis, orthogonal factor analysis, classification, and clustering techniques--are tailored and combined into a system designed to digest high-dimensional vectors of data on operational readiness of Navy ships. Such data consist of large numbers of scores for individual ships assigned by experts. The purpose of the data reduction system is to provide a robust method of representing the data by a small number of scores that are meaningfully related to the original scores and that allow classification and clustering of the ships into homogeneous groups on relevant readiness scales. Simulated data drawn from mixtures of specified multivariate normal populations have been used to test the ability of the system to recover individual populations and to detect trends over time.

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

READINESS EVALUATIONS USING MULTIVARIATE DATA REDUCTION

by

Zeev Barzily
W. H. Marlow
S. Zacks

1. Introduction and Summary

Readiness evaluation is one of the most important problem areas in the study of complex military systems. Such studies usually encompass a large number of measurements on the performance of the various subsystems and then attempt to construct reasonable models that relate the evaluation indices of the subsystems in a meaningful functional manner. This is indeed very often a formidable task due to the complexity and multiplicity of variables and functions. However, it is often the case that many of the measured variables correlate with each other. These intercorrelations reveal that variables contain some information on each other. Accordingly, if these intercorrelations can be utilized in a manner that allows considerable reduction in the number of factors to be considered, without much loss in the information in the original data, a significant step can be taken towards simplification of the problem. The present paper applies several well-known multivariate statistical methods to attain this goal. The main objective of the present paper is to discuss what some of the available multivariate statistical methods can attain and to show that such methods can be easily implemented by utilizing appropriate computer packages. In particular, we discuss the methods of principal and rotated factor analysis, and discriminant analysis. We apply these methods to simulated data on 21 operational readiness variables related to Navy destroyers. The variables and the corresponding parameters were taken from the Institute of Naval Studies study [5]. This study analyzes actual data collected over several years on 83 destroyers. It extends to various aspects of the readiness problem and relates operational readiness to material readiness. As mentioned in our recent survey paper [3], we believe that study [5] is of fundamental importance. It employs a variety of multivariate statistical procedures in a penetrating manner and provides a sound analysis of complicated problems. Our intention is not to duplicate that study but to provide an exposition on the application of the multivariate methods mentioned above. We have chosen to create data sets by simulation and not to use actual data since in this way we can generate data following multivariate normal distributions having specific structures. Thus, by applying the multivariate methods on different sets of simulated data we can illustrate the strength of the methods and what can actually be achieved. We will show that the systems (destroyers) in this example can be classified according to the values of two or three factor scores, which relate all 21 variables in an orthogonal fashion. The factor scores can be graphed and their periodic determination can provide important follow-up on the state of readiness. Statistical control charts can be devised to provide early detection of deterioration in the state of readiness. Similarly, if the data consists of a mixture of two or more samples from different multivariate populations, the plotting of factor scores obtained by a factor-analysis of the whole data set can reveal the existence of different clusters. These ideas will be demonstrated in the present paper. We start in Section 2 with a description of the simulations and the structure of the data sets. Section 3 is devoted to principal and rotated factor analysis. In Section 4 we discuss the application of factor analysis to detecting changes in the state of readiness of systems. Here we also compare discrimination and classification based on the factor scores to those obtained by step-wise discriminant analysis. The mathematical development is presented in appendices. We also provide some computer programs.

2. Simulating Data Sets

In the present study we construct data sets on the basis of the operational readiness indices, ORI, of the following 21 variables.

v_1	Ship control	SHC
v_2	Navigation	NAV
v_3	Surface operations - CIC (Combat Information Center)	SOPS
v_4	Battle communications	BATC
v_5	Surface gunnery (non-firing)	SGUN
v_6	AAW (Anti-air Warfare) - CIC	AAWC
v_7	AAW - Weapons control	AAWN
v_8	Engineering	ENG
v_9	Setting material condition	SMC
v_{10}	Damage control	DC
v_{11}	NBC (Nuclear, biological, and chemical)	NBC
v_{12}	Low-visibility piloting	LVP
v_{13}	CIC - Assistance in piloting	CICAP
v_{14}	CIC - Assistance in ASW (Anti-submarine warfare)	CICASW
v_{15}	ECM (Electronic countermeasures)	ECM
v_{16}	Modified full-power run	BFPR
v_{17}	Surface firing	SFIR
v_{18}	AA firing	AAF
v_{19}	Gunfire support	GUNS
v_{20}	Communications	COMM
v_{21}	ASW operations	ASW

The raw scores obtained on these variables by the 83 ships during training can be obtained in the Institute of Naval Studies [5]. We consider rather the ORI's which are indices obtained from the raw scores by the transformation

$$ORI = 5 + 2\ (NSCORE) \qquad (2.1)$$

where NSCORE denotes the standard normal fractile corresponding to the percentile point of the raw score. More precisely, if $x_{(1)} \leqq \cdots \leqq x_{(n)}$ is the order statistic of a sample of n observations on a variable x, the NSCORE corresponding to $x_{(i)}$ is $z_{(i)} = \Phi^{-1}\left(\frac{i}{n+1}\right)$, $i = 1,\ldots,n$; where $\Phi(z)$ is the standard normal C.D.F. Theoretically the ORI values, of each of the variables v_i $(i=1,\ldots,21)$, are normally distributed with mean $E\{v_i\} = 5$ $(i=1,\ldots 21)$ and variance $Var\{v_i\} = 4$ $(i=1,\ldots 21)$. In addition, the ORI variables, v_i, are not independent. We assume that the vector $\underset{\sim}{v} = [v_1,\ldots,v_{21}]^T$ has a multinormal distribution $N(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$, with mean vector $\underset{\sim}{\mu} = 5\underset{\sim}{1}_{21}$, where $\underset{\sim}{1} = [1,\ldots,1]^T$, and covariance matrix $\underset{\sim}{\Sigma} = 4\underset{\sim}{R}$ where $\underset{\sim}{R}$ denotes the matrix of intercorrelations among the 21 variables v_i. For the purpose of simulating data sets we have used the matrix $\underset{\sim}{R}$ given in the Institute of Naval Studies [5] and presented here in Table 1. The simulation was performed according to an algorithm described in Appendix I. It is based on simulating independent standard normal variates, z, and transforming them to corresponding v_i variates $(i=1,\ldots,21)$ by employing recursive relations between joint and conditional multinormal distributions. A Fortran program for such simulation is given in Appendix IV. In Table 2 we present a sample of $n = 50$ vectors of six variables $(v_1, v_5, v_6, v_7, v_{12}, v_{14})$ simulated according to this program. The sample means, standard deviations and intercorrelations are provided in Table 3. As illustrated, the sample statistics are generally deviating to some extent (according to their sampling distributions) from the parameters used. However, in actual cases the population parameters are unknown and the analysis must be based exclusively on the sample values, with the possible incorporation of some prior information, and this is what we are doing here.

TABLE 1

INTERCORRELATIONS BETWEEN ORI VARIABLES v_1 to v_{21}

1	1.00	0.17	0.27	0.04	0.38	0.31	0.45	0.10	0.06	0.03	
	0.08	0.32	0.36	0.06	0.17	0.16	0.07	0.12	0.03	0.01	0.10
2	0.17	1.00	0.22	0.06	0.18	0.03	0.14	0.22	0.01	0.20	
	0.04	0.12	0.14	0.26	0.01	0.07	0.02	0.09	0.09	0.03	0.13
3	0.27	0.22	1.00	0.06	0.18	0.03	0.14	0.22	0.01	0.20	
	0.04	0.12	0.14	0.26	0.01	0.07	0.02	0.09	0.09	0.03	0.13
4	0.04	0.06	0.09	1.00	0.23	0.23	0.13	0.11	0.13	0.10	
	0.08	0.12	0.35	0.03	0.19	0.25	0.03	0.10	0.03	0.04	0.06
5	0.38	0.18	0.23	0.28	1.00	0.34	0.05	0.08	0.04	0.01	
	0.12	0.07	0.03	0.04	0.04	0.25	0.10	0.20	0.03	0.18	0.02
6	0.31	0.03	0.23	0.34	0.27	1.00	0.29	0.13	0.13	0.03	
	0.05	0.03	0.22	0.15	0.06	0.17	0.12	0.13	0.04	0.03	0.10
7	0.45	0.14	0.13	0.05	0.29	0.36	1.00	0.17	0.17	0.03	
	0.09	0.14	0.17	0.11	0.17	0.14	0.01	0.31	0.01	0.12	0.08
8	0.10	0.22	0.11	0.08	0.13	0.17	0.06	1.00	0.09	0.10	
	0.00	0.21	0.40	0.01	0.03	0.10	0.17	0.28	0.12	0.07	0.08
9	0.06	0.01	0.13	0.04	0.13	0.17	0.09	0.33	1.00	0.18	
	0.19	0.07	0.14	0.04	0.14	0.21	0.19	0.05	0.07	0.13	0.18
10	0.03	0.20	0.10	0.01	0.03	0.03	0.10	0.18	0.48	1.00	
	0.68	0.06	0.04	0.07	0.20	0.17	0.03	0.02	0.02	0.29	0.18
11	0.08	0.04	0.08	0.17	0.05	0.09	0.00	0.19	0.68	0.50	
	1.00	0.11	0.02	0.11	0.03	0.09	0.07	0.04	0.13	0.06	0.13
12	0.32	0.12	0.12	0.07	0.03	0.14	0.21	0.07	0.06	0.11	
	0.08	1.00	0.00	0.06	0.18	0.25	0.07	0.07	0.01	0.23	0.21
13	0.36	0.14	0.35	0.03	0.22	0.17	0.40	0.14	0.04	0.02	
	0.00	0.55	1.00	0.12	0.04	0.20	0.11	0.08	0.20	0.17	0.05
14	0.06	0.26	0.03	0.04	0.15	0.11	0.01	0.04	0.07	0.11	
	0.06	0.12	0.19	1.00	0.16	0.29	0.04	0.03	0.06	0.06	0.05
15	0.17	0.01	0.19	0.04	0.06	0.17	0.03	0.14	0.20	0.03	
	0.18	0.04	0.16	0.10	1.00	0.13	0.14	0.14	0.02	0.34	0.51
16	0.16	0.07	0.25	0.25	0.17	0.14	0.10	0.21	0.17	0.09	
	0.25	0.20	0.29	0.13	0.13	1.00	0.01	0.14	0.09	0.28	0.13
17	0.07	0.02	0.03	0.10	0.12	0.01	0.17	0.19	0.03	0.07	
	0.07	0.11	0.04	0.14	0.01	0.09	1.00	0.00	0.05	0.24	0.23
18	0.12	0.09	0.10	0.20	0.13	0.31	0.28	0.05	0.02	0.04	
	0.07	0.08	0.03	0.14	0.14	0.00	0.10	1.00	0.11	0.04	0.12
19	0.03	0.09	0.03	0.03	0.04	0.01	0.12	0.07	0.02	0.13	
	0.01	0.20	0.06	0.02	0.09	0.05	0.11	0.07	1.00	0.12	0.03
20	0.01	0.03	0.04	0.18	0.03	0.12	0.07	0.13	0.29	0.06	
	0.23	0.17	0.06	0.34	0.28	0.24	0.04	0.12	0.17	1.00	0.03
21	0.10	0.13	0.06	0.02	0.10	0.08	0.08	0.18	0.18	0.13	
	0.21	0.05	0.05	0.51	0.13	0.23	0.12	0.03	0.03	0.09	1.00

TABLE 2

50 SIMULATED VECTORS OF SIX VARIABLES
ACCORDING TO THE MULTINORMAL DISTRIBUTION $N(5\underset{\sim}{1}, 4\underset{\sim}{R})$

i	v_1	v_5	v_6	v_7	v_{12}	v_{14}
1	8.508	8.269	7/591	7.124	8.463	5.613
2	4.529	5.294	1.966	5.417	4.021	6.608
3	5.594	5.933	0.847	5.340	6.141	4.522
4	1.891	2.948	7.371	5.303	5.938	5.051
5	2.681	1.105	2.656	1.402	4.078	4.125
6	6.837	4.298	5.896	6.648	6.902	7.504
7	4.968	6.267	4.629	5.431	5.985	2.484
8	2.956	2.448	1.450	1.762	3.117	5.060
9	8.007	3.538	4.084	4.136	4.800	4.747
10	3.961	3.451	3.985	3.774	1.771	3.184
11	5.949	7.807	3.955	6.362	3.873	6.348
12	4.983	4.908	3.409	7.550	4.702	6.852
13	4.244	4.284	2.781	4.367	5.851	4.362
14	3.873	3.360	1.704	4.349	4.340	6.003
15	5.455	2.531	9.759	4.765	4.340	5.939
16	6.076	6.838	6.972	2.700	5.236	7.398
17	2.430	6.347	7.090	1.482	3.639	-.140
18	4.616	3.585	5.253	7.105	5.440	3.175
19	3.897	4.281	3.646	5.297	5.004	7.421
20	7.915	5.708	6.189	8.844	6.734	10.158
21	4.339	4.453	0.926	2.260	3.118	5.672
22	9.427	6.974	9.020	7.265	7.532	4.959
23	6.122	5.779	5.043	6.474	5.405	8.099
24	7.042	6.516	2.866	7.566	3.579	6.980
25	4.872	6.078	9.149	5.884	2.773	4.261
26	5.922	5.126	4.973	5.062	3.671	3.683
27	2.836	3.512	5.607	0.082	3.345	1.516
28	4.626	3.806	2.232	4.252	7.386	6.028
29	4.007	3.138	5.365	1.818	4.890	1.473
30	6.566	7.718	8.269	7.856	6.988	3.450
31	5.278	3.930	5.021	2.825	2.912	2.715
32	8.079	4.857	3.980	5.727	8.746	7.437
33	4.797	3.643	5.791	5.551	9.067	2.724
34	1.587	-.807	3.601	3.556	5.196	2.714
35	1.722	2.332	3.334	1.992	1.319	0.916
36	2.967	3.897	1.683	1.195	5.045	2.528
37	4.441	3.769	6.153	1.230	6.165	4.472
38	5.537	5.391	5.139	3.380	0.530	6.516
39	2.584	4.352	5.634	3.576	5.014	1.543
40	5.141	5.169	2.510	2.849	4.054	4.894
41	5.726	2.640	6.141	10.725	8.585	11.800
42	4.648	5.487	5.880	6.518	5.053	7.469
43	7.128	3.847	3.674	3.455	5.580	4.176
44	5.108	7.783	2.857	3.857	4.193	2.878
45	6.176	5.253	5.429	6.230	6.403	3.548
46	0.606	1.079	5.370	3.985	2.871	2.396
47	7.174	4.692	3.537	6.435	4.326	9.936
48	0.549	6.349	5.689	3.537	6.540	2.525
49	3.342	1.917	5.073	4.799	2.103	6.821
50	5.230	6.934	1.036	7.321	6.904	5.135

TABLE 3

SAMPLE STATISTICS OF
THE SIX VARIABLES IN TABLE 2

SAMPLE MEANS AND STAND.DEV.

	v_1	v_5	v_6	v_7	v_{12}	v_{14}
MEANS	4.860	4.577	4.644	4.728	4.993	4.914
ST. DEVS.	1.998	1.897	2.180	2.244	1.916	2.468

CORRELATIONS MATRIX

	v_1	v_5	v_6	v_7	v_{12}	v_{14}
v_1	1.000000	0.535377	0.183915	0.558162	0.399037	0.519481
v_5	0.535377	1.000000	0.158194	0.362251	0.238443	0.128076
v_6	0.183915	0.158194	1.000000	0.237437	0.192687	-.023431
v_7	0.558162	0.362251	0.237437	1.000000	0.492125	0.617677
v_{12}	0.399037	0.238443	0.192687	0.492125	1.000000	0.239948
v_{14}	0.519481	0.128076	-.023431	0.617677	0.239948	1.000000

It should also be remarked that the simulation is based on the matrix of intercorrelations of the above six variables only. This matrix is, however, a submatrix of that given in Table 1 and can be obtained by reading the appropriate rows and columns. In the course of the present study several different data sets were simulated, employing the same algorithm with only slight modifications from case to case, as will be explained later.

3. Principal and Rotated Factor Analysis

It is generally difficult to make comprehensive inference of multivariate data without further analysis, due to the large number of intercorrelated variables. Even in the case of only six variables it will be difficult to discriminate between "good" and "bad" systems, just by inspecting the data sets, or by performing a univariate analysis on each variable separately. The methods of multivariate analysis are designed to provide the needed information in cases of many variables which are highly correlated. In the present section we discuss the methods of principal and rotated factor analysis, and show how they can be applied to the evaluation of the readiness of systems. An outline of the theory is given in Appendix II. We refer the reader for an extensive development of the theory and computer programs to the books of Overall and Klett [6], Cooley and Lohnes [4], Tatsuoka [8], and Van de Geer [9].

3.1 Principal Factor Analysis

The main objective of principal factor analysis is to provide a small number, m , of linear combinations of the original variables $v_1,\ldots,v_p$ $(2\leq m<p)$ so that (i) a large proportion of the total variance of the original variables should be accounted for by the m transformed variables, and (ii) the transformed variables should be uncorrelated. It is shown in Appendix II that the solution of this problem is obtained by determining first the m largest eigenvalues of $\underset{\sim}{R}$ and the corresponding eigenvectors; followed by determination of factor scores for each system. Let $\lambda_1 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of the $p \times p$ correlation matrix $\underset{\sim}{R}$. Since $\underset{\sim}{R}$ is positive definite, these eigenvalues are all real and positive (with probability one). Moreover $\lambda_1+\cdots+\lambda_p = p$. Hence, choose m so that $(\lambda_1+\cdots+\lambda_m)/p$ is "close enough" to 1. This ratio is the proportion of the sum of variances of $v_i(i=1,\ldots,p)$ that is accounted for (explained) by the m factors. These factors are constructed in the following manner. Let $\underset{\sim}{b}_{(j)}$ $(j=1,\ldots,m)$ be the orthonormal eigenvector of $\underset{\sim}{R}$ corresponding to $\lambda_j(j=1,\ldots,m)$. The m factor-score variables corresponding to $\underset{\sim}{v} = [v_1,\ldots,v_p]^T$ are given by

$$f_j = \frac{1}{\sqrt{\lambda_j}} \underset{\sim}{b}_{(j)}^T \underset{\sim}{u} , \qquad j = 1,\ldots,m \tag{3.1}$$

where $\underset{\sim}{u} = [u_1,\ldots,u_p]^T$ is a vector of standard scores corresponding to $\underset{\sim}{v}$, i.e., $u_i = (v_i-\bar{v}_i)/s_i$, $i = 1,\ldots,p$, $\bar{v}_i$ denotes the sample mean of the ith variable and s_i designates its sample standard deviation. The m factor-scores are computed for each unit in the sample. The sample factor-scores are uncorrelated standard scores. The coefficients b_{ij} in (3.1) $(i=1,\ldots,p)$ are sometimes interpretable in terms of the original ORI variables. We illustrate these ideas first on the $p = 6$ ORI scores of the sample of 50 units given in Table 2. (Further analysis of all 21 ORI variables appears below.) The numerical results presented were obtained by Program SIMU, presented in Appendix IV. The eigenvalues and eigenvectors of the correlation matrix presented in Table 3 are given in Table 4. These values were computed by employing a computer library subroutine which determines the eigenvalues and eigenvectors of a symmetric matrix. This subroutine program was merged into Program SIMU.

TABLE 4

EIGENVALUES AND EIGENVECTORS OF THE CORRELATION MATRIX IN TABLE 3

EIGENVALUES OF CORREL. MATRIX

λ_5	λ_3	λ_2	λ_6	λ_4	λ_1
0.384321	0.858006	1.068542	0.239473	0.700426	2.749232

EIGENVECTORS

$b_{(5)}$	$b_{(3)}$	$b_{(2)}$	$b_{(6)}$	$b_{(4)}$	$b_{(1)}$
0.707698	0.266720	-.035902	0.399522	-.115832	0.503687
-.329692	0.749462	0.292523	-.336753	0.026661	0.360454
0.072605	-.410358	0.755713	-.156839	-.440037	0.192362
-.608214	-.220703	-.118198	0.546450	-.067590	0.514023
0.117372	-.321752	0.141075	-.231162	0.312732	0.335783
0.038669	-.215730	-.555123	-.591394	-.356564	0.407326

If we consider $m=3$ factor-scores in Table 4, we see that the proportion of explained variability is $\frac{1}{6}(\lambda_1+\lambda_2+\lambda_3) = .78$. (In other words, three factor-scores out of six explain close to 80% of the correlation matrix $\underset{\sim}{R}$.) Notice that the first factor-score, f_1, weights positively each one of the six ORI variables, v_1, v_5, v_6, v_7, v_{12}, v_{14}. However, variables v_1 and v_7 obtain weights which are about 1.22 times larger than those of v_5, v_{12} and v_{14} and about 2.6 times larger than that of v_6. In terms of the ship functions, factor f_1 emphasizes Ship control and AAW-Weapons control considerably more than AAW-CIC or Low-visibility piloting, etc. The second factor-score, f_2, emphasizes v_6 (AAW-CIC) and gives a large negative weight to v_{14}. The third factor-score emphasizes v_5 (Surface gunnery) and deemphasizes v_6, v_7, v_{12}, v_{14}. Thus, the three dimensions of this factor analysis are: "Ship and AAW control," "Radar and information communication" and "Surface gunnery." In Table 5 we provide the three factor-scores of the 50 units in the sample. Figure 1 presents a scattergram of f_2 versus f_1. By indicating on the scattergram the unit number of each point we can immediately discriminate between systems having high factor-scores on the two dimensions (like 30,22,2) and those having low factor-scores on both dimensions (like 8). Furthermore, by indicating the zero lines (solid) and the lines at ± 1 (broken) we can obtain further information on the state of readiness of the systems in the sample. For example, systems 47 and 41 have high scores on f_1 but are in the lower tail of f_2. This means that their control functions are "good" but the CIC functions are "bad." Similarly, systems 27 and 17 are low on f_1 but high on f_2. Although this information is given also in Table 2, it is often confounded and obscured. A scattergram of the factor-scores provides a convenient expression of the relative state of readiness of the systems.

Suppose now that the sample of 50 systems consists of two subsamples from two different multinormal populations. What will the factor scores reveal? To illustrate numerically the result of principal factor analysis on such a mixture of samples we simulated a data set in which the first 25

TABLE 5

FACTOR-SCORES OF DATA IN TABLE 2

FACTOR SCORES

	f_1	f_2	f_3
1	1.9566	1.4487	0.5523
2	0.0350	-1.2588	0.7455
3	0.2499	-.9482	1.2197
4	-.2853	0.7313	-1.9226
5	-1.4842	-.8711	-.7976
6	1.0904	-.1818	-.8829
7	0.1847	0.8086	0.7144
8	-1.3267	-1.3706	0.1078
9	0.2075	-.3450	0.2880
10	-.9964	-.1781	0.3727
11	0.7315	-.2433	1.5691
12	0.5385	-.9534	-.0196
13	-.2271	-.4586	0.0996
14	-.4693	-1.4149	-.0079
15	0.1561	1.1283	-1.8089
16	0.5649	0.6771	0.6035
17	-1.1530	2.2953	0.9748
18	0.0907	0.3495	-.7513
19	0.0963	-.9362	-.3611
20	1.9788	-.5931	-.6389
21	-.7845	-1.4294	1.1580
22	1.8651	1.7375	0.0577
23	0.9594	-.4617	0.0530
24	0.8856	-1.0400	1.2631
25	0.2385	1.6594	0.0677
26	0.0046	0.3300	0.6407
27	-1.5593	1.0576	0.1709
28	0.0835	-.9679	-.3610
29	-1.0137	0.9316	-.2311
30	1.3408	1.9558	0.2934
31	-.7254	0.4495	0.4945
32	1.3312	-.5698	-.3064
33	0.3345	0.9712	-1.2602
34	-1.5269	-.5433	-2.2614
35	-2.0267	0.0275	0.1903
36	-1.2431	-.3590	0.6296
37	-.4610	0.7503	-.5110
38	-.3461	-.3224	1.1453
39	-.8118	1.1315	-.1882
40	-.3786	-.5993	1.0984
41	1.9401	-1.3501	-2.9433
42	0.6470	-.0890	-.3354
43	0.0315	-.2067	0.3112
44	-.1105	0.3046	2.1961
45	0.5618	0.6622	0.0322
46	-1.6204	0.2301	-1.5515
47	0.9613	-1.6218	0.0733
48	-.6117	1.3302	-.0064
49	-.6640	-.8513	-1.1043
50	0.7599	-.9025	1.1780

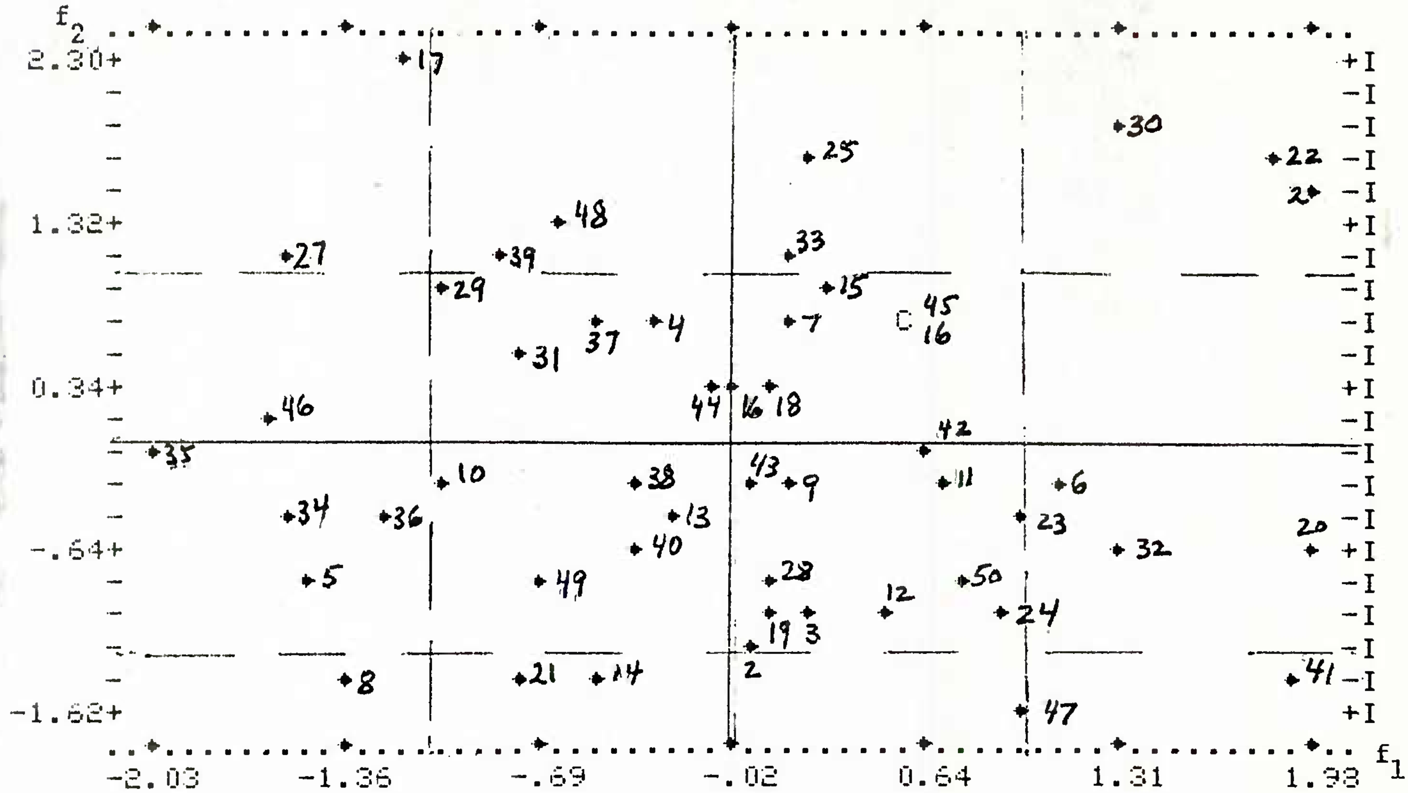


Figure 1. Scattergram of the Factor-Scores f_2 Versus f_1 for the Data of Table 5.

vectors are identical with the first 25 lines of Table 2 $(N(5\underset{\sim}{1}, 4\underset{\sim}{R}))$ and the last 25 vectors are simulated from $N(\underset{\sim}{1}, 4\underset{\sim}{R})$. The sample means, standard deviations, correlation matrix, eigenvalues and eigenvectors of these 50 vectors are given in Table 6. The corresponding three factor-scores are given in Table 7. We see in Table 6 that in such a mixture the correlations become larger and so is the maximal eigenvalue. The three factor-scores account here for 90% of the variability reflected in $\underset{\sim}{R}$. In Figure 2 we provide a scattergram of f_2 versus f_1 . The points

TABLE 6

SAMPLE STATISTICS OF A SAMPLE OF 50 UNITS CONSISTING OF 25 UNITS FROM $N(5\underset{\sim}{1},4\underset{\sim}{R})$ AND 25 UNITS FROM $N(\underset{\sim}{1},4\underset{\sim}{R})$

SAMPLE MEANS AND STAND.DEV.

	v_1	v_5	v_6	v_7	v_{12}	v_{14}
MEANS	2.860	2.577	2.644	2.728	2.993	2.914
ST. DEVS.	3.107	3.013	3.030	3.289	2.723	3.519

CORRELATIONS MATRIX

1.000000	0.811830	0.621012	0.804749	0.736423	0.780947
0.811830	1.000000	0.614419	0.723784	0.667859	0.611753
0.621012	0.614419	1.000000	0.628804	0.592935	0.490986
0.804749	0.723784	0.628804	1.000000	0.768023	0.817156
0.736423	0.667859	0.592935	0.768023	1.000000	0.642159
0.780947	0.611753	0.490986	0.817156	0.642159	1.000000

EIGENVALUES OF CORREL. MATRIX

λ_5	λ_3	λ_2	λ_1	λ_4	λ_6
0.167197	0.368039	0.550869	4.455704	0.336945	0.121246

EIGENVECTORS

$\underset{\sim}{b}_{(5)}$	$\underset{\sim}{b}_{(3)}$	$\underset{\sim}{b}_{(2)}$	$\underset{\sim}{b}_{(1)}$	$\underset{\sim}{b}_{(4)}$	$\underset{\sim}{b}_{(6)}$
0.610051	0.276820	-.119073	0.438211	-.141319	-.570114
-.288832	0.764794	0.158765	0.406998	-.090066	0.364282
0.080816	-.390887	0.796742	0.356930	-.270797	0.071749
-.698836	-.227609	-.194875	0.436857	-.004906	-.480602
0.159404	-.147009	0.007566	0.404763	0.366899	0.193840
0.155249	-.335038	-.536455	0.400310	-.383467	0.518236

TABLE 7

FACTOR-SCORES OF 50 SYSTEMS IN MIXED SAMPLE

FACTOR SCORES

	f_1	f_2	f_3
1	1.8249	0.9802	0.7472
2	0.6881	-1.1033	0.5479
3	0.7699	-1.0667	1.3573
4	0.7072	1.1176	-1.8837
5	-.0472	-.2299	-.7801
6	1.3266	-.1723	-.9027
7	0.3453	0.7399	0.9243
8	-.0050	-.8006	-.0236
9	0.8003	-.1696	0.2430
10	0.1986	0.3366	0.1892
11	1.0899	-.3159	1.3313
12	0.9697	-.8607	-.1956
13	0.5917	-.3196	0.2196
14	0.4286	-1.0886	-.1133
15	0.9535	1.6044	-1.9780
16	1.1269	0.7595	0.4347
17	0.2631	2.5942	1.1327
18	0.7891	0.5113	-.6330
19	0.7804	-.7011	-.5271
20	1.7745	-.7450	-.8689
21	0.2509	-1.0806	0.9768
22	1.7912	1.4666	0.2062
23	1.2417	-.4463	-.1469
24	1.1088	-1.0769	0.9729
25	0.9771	1.9196	-.1751
26	-.7617	0.1372	0.6030
27	-1.4890	1.2482	0.2769
28	-.7488	-1.2636	-.1629
29	-1.2324	0.9516	0.0151
30	0.0280	1.2930	0.5041
31	-1.1250	0.4772	0.4726
32	-.0888	-1.1502	-.0981
33	-.5271	0.5583	-.7421
34	-1.5472	-.2213	-2.0362
35	-1.8205	0.3797	0.1596
36	-1.4303	-.4117	0.8546
37	-.9047	0.6886	-.3415
38	-.9355	-.2245	0.6488
39	-1.1117	1.0521	0.0347
40	-.9956	-.7624	1.0585
41	0.2710	-1.7159	-3.0711
42	-.3801	-.3385	-.5088
43	-.7755	-.4418	0.4576
44	-.8467	-.1183	2.2624
45	-.4526	0.2396	0.2413
46	-1.5472	0.6127	-1.5592
47	-.3164	-1.8582	-.2855
48	-.9589	1.1256	0.2738
49	-1.0917	-.5511	-1.4628
50	-.4574	-1.5583	1.3502

in Figure 2 representing systems 26-50 have been circled. We see a very clear separation between the two subsamples. Although this example may seem somewhat artificial it shows that principal factor analysis may reveal clustering of subsamples in case the populations are sufficiently distinct. A discriminant analysis can then be performed.

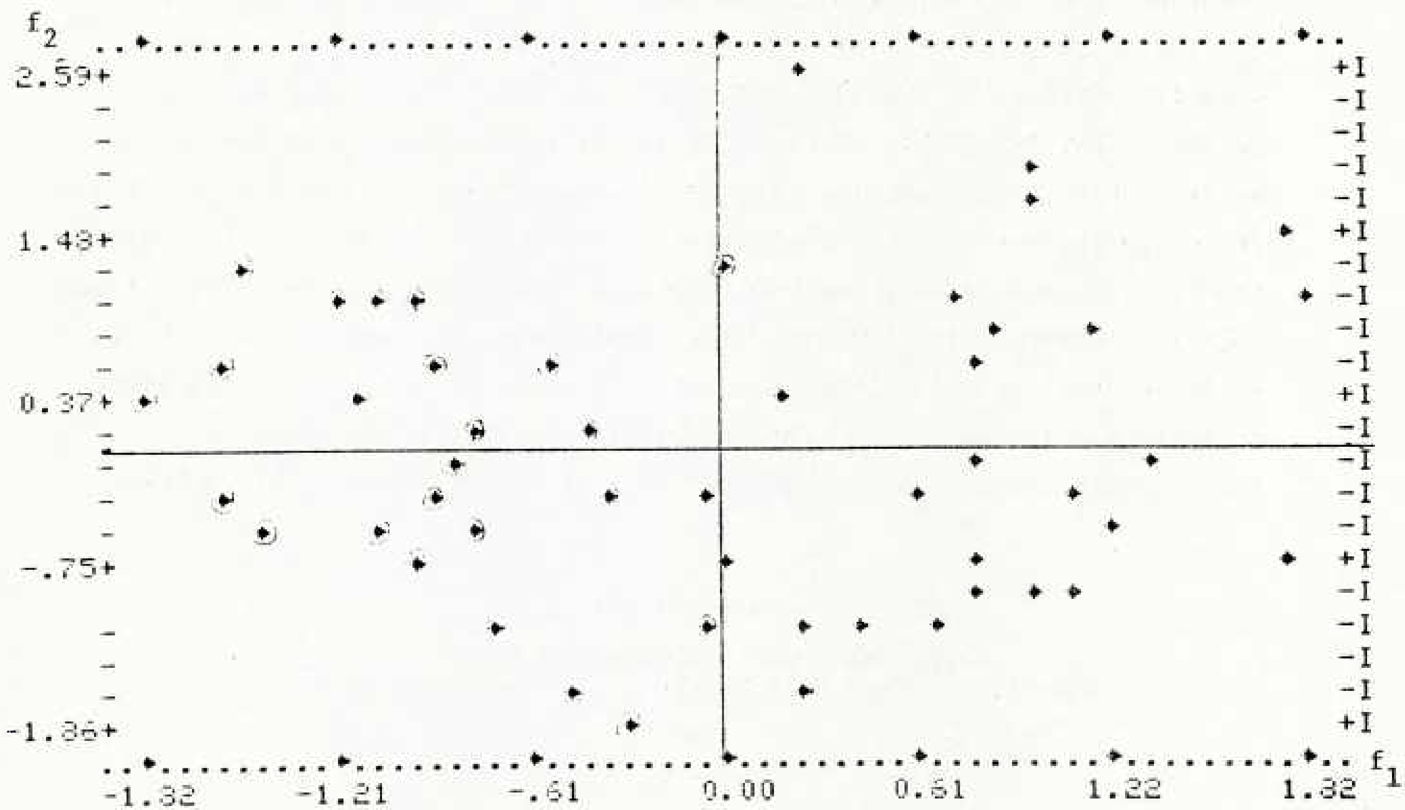


Figure 2. Scattergram of f_2 Versus f_1 in Mixed Sample.

3.2 Rotated (Orthogonal) Factor Analysis

In the present section we consider the method of factor analysis, when the number of variables (21) is considerably larger than the number of variables (6) in the examples of the previous section. The simulation is based on Program SNOR2, which simulates (raw) ORI scores and writes them directly into a data file "RSCORE" in the computer. The sample means, standard deviations are written into a file "SMNS" and the intercorrelations are written into a computer file "MATRIX". The files "RSCORE" and "SMNS" are then read into a program which computes the standardized variables z_i $(i=1,\ldots,21)$. These standardized values are written into a computer data file "ZSCORE". The files "MATRIX" and "ZSCORE" can then be used to perform a rotated factor analysis by one of the computer programs which may be available in the computer library. We have chosen to use the SPSS [7] factor analysis program. The program has the option of reading the correlation matrix, rather than the whole data set. We therefore merged the file "MATRIX" into the SPSS program. The sample means and standard deviations of the sample of 21 variables simulated according to $N(5\underset{\sim}{1}, 4\underset{\sim}{R})$ are given in Table 8. The matrix $\underset{\sim}{R}$ for the simulation is that of Table 1. The sample correlation matrix is of size 21 x 21 and will not be given here.

TABLE 8

SAMPLE MEANS AND STANDARD DEVIATIONS OF SIMULATED SAMPLE OF 50 VECTORS OF 21 ORI VARIABLES

	4.755	5.277	4.913	5.450	4.897	4.753	4.891
MEANS	5.065	4.567	5.047	4.573	4.397	4.737	5.378
	4.666	4.952	4.468	4.941	4.697	5.012	5.080
	2.023	2.323	2.133	1.932	1.967	2.358	2.186
ST. DEVS.	2.099	1.752	2.218	1.723	1.794	2.120	2.388
	1.918	2.177	1.515	2.263	1.759	2.043	2.628

The first step in the SPSS program is to compute the eigenvalues and eigenvectors of the correlation matrix. All the eigenvalues are given in Table 9.

TABLE 9

EIGENVALUES OF THE SAMPLE CORRELATION MATRIX

FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
1	4.71774	22.5	22.5
2	3.04975	14.5	37.0
3	2.15763	10.3	47.3
4	1.71361	8.2	55.4
5	1.45501	6.9	62.4
6	1.22593	5.8	68.2
7	1.11568	5.3	73.5
8	.95484	4.5	78.0
9	.77374	3.7	81.7
10	.69166	3.3	85.0
11	.60214	2.9	87.9
12	.52346	2.5	90.4
13	.40730	1.9	92.3
14	.32128	1.5	93.9
15	.29829	1.4	95.3
16	.25938	1.2	96.5
17	.23882	1.1	97.6
18	.16924	.8	98.5
19	.13962	.7	99.1
20	.11074	.5	99.6
21	.07412	.4	100.0

The eigenvectors of $\underset{\sim}{R}$ are given in a matrix $\underset{\sim}{S}$ of size $p \times m$, called the <u>factor (structure) matrix</u>. The m column vectors of $\underset{\sim}{S}$ are defined as

$$\underset{\sim}{s}_{(j)} = \lambda_j^{1/2} \, \underset{\sim}{b}_{(j)} , \qquad j = 1,\ldots,m \qquad (3.2)$$

where $\underset{\sim}{b}_{(j)}$ is the eigenvector corresponding to λ_j , $\lambda_1 \geqq \cdots \geqq \lambda_m > 0$, the m largest eigenvalues of $\underset{\sim}{R}$. Obviously $|| \underset{\sim}{s}_j ||^2 = \lambda_j$. Furthermore, if $m = p$ then $\underset{\sim}{R} = \sum_{j=1}^{p} \underset{\sim}{s}_{(j)} \underset{\sim}{s}_{(j)}^T$ is the spectral decomposition. Let $\hat{\underset{\sim}{R}}_m = \sum_{j=1}^{m} \underset{\sim}{s}_j \underset{\sim}{s}_j^T$ and $\tilde{\underset{\sim}{R}}_m = \underset{\sim}{R} - \hat{\underset{\sim}{R}}_m$. It is desirable to choose m so that $\tilde{\underset{\sim}{R}}_m$ is negligible (or statistically non-significant). Tests of the significance of $\tilde{\underset{\sim}{R}}_m$ are available (see Cooley and Lohnes [4,103]. According to the spectral decomposition,

$$h_i = \sum_{j=1}^{m} \lambda_j b_{ij}^2 , \qquad i = 1,\ldots,p \tag{3.3}$$

is the part of the ith diagonal element of $\underset{\sim}{R}$ (the variance of z_i) which is explained by $\underset{\sim}{S}$. The parameter h_i is called the communality of the ith variable. In Table 10 we present the factor (structure) matrix and the communalities.

The factor-scores defined in (31) can be obtained as inner products of $\underset{\sim}{z}$ with the column vectors of $\underset{\sim}{F} = \underset{\sim}{S} \underset{\sim}{\Lambda}^{-1}$, where $\underset{\sim}{\Lambda}$ is an m x m diagonal matrix with entries $\lambda_1,\ldots,\lambda_m$. When the number of variables is large, it is generally difficult to interpret the factor-scores obtained by the matrix $\underset{\sim}{F}$ (see Table 10). For this reason various rotation techniques were developed, which transform the factor (structure) matrix $\underset{\sim}{S}$ to a matrix $\underset{\sim}{A} = \underset{\sim}{S}\underset{\sim}{P}$, $\underset{\sim}{P}$ being an m x m (orthogonal) matrix so that the column vectors of $\underset{\sim}{A}$ have as many zero entries as possible. We consider here an orthogonal rotation matrix $\underset{\sim}{P}$ obtained by a method called <u>varimax</u>. This method maximizes the variance of the coefficients of each column vector of $\underset{\sim}{A}$ (see Van de Geer [9,150]). In Table 11 we present the varimax rotation of the matrix $\underset{\sim}{S}$ of Table 10, as obtained by the SPSS program.

TABLE 10

THE FACTOR (STRUCTURE) MATRIX $\underset{\sim}{S}$ AND COMMUNALITIES

	FACTOR 1	FACTOR 2	FACTOR 3	COMMUNALITY
V1	.63656	-.38894	-.08892	.56439
V2	.40868	.48347	-.46057	.61289
V3	.43219	-.09805	-.36006	.32604
V4	.39535	-.13877	.46957	.39606
V5	.55830	-.22559	-.16417	.38953
V6	.55146	-.58565	.30006	.73713
V7	.60517	-.41661	-.20766	.58292
V8	.45170	.54153	-.22955	.54998
V9	.43472	.52352	.41075	.63178
V10	.23966	.70879	-.11495	.57303
V11	.49398	.42669	.52152	.69808
V12	.43682	-.21725	-.16080	.26386
V13	.58726	-.30366	-.39032	.58943
V14	.54332	.20345	.07008	.34150
V15	.47193	-.06378	.37983	.37105
V16	.58146	.11887	-.23283	.40643
V17	.39548	-.08428	-.40533	.32780
V18	.42930	-.40516	.36584	.48229
V19	.09042	.47153	-.13407	.24849
V20	.39481	.19576	.47195	.41694
V21	.47903	.43095	.01722	.41549

TABLE 11

VARIMAX ROTATED FACTOR MATRIX $\underset{\sim}{A}$

	FACTOR 1	FACTOR 2	FACTOR 3
V1	.69162	-.09886	.27619
V2	.31604	.70436	-.12995
V3	.54405	.16496	-.05324
V4	.12286	-.11620	.60618
V5	.59764	.04061	.17524
V6	.52327	-.41322	.54090
V7	.73965	-.10143	.15984
V8	.20729	.70680	.08628
V9	-.11529	.50582	.60217
V10	-.08217	.74767	.06537
V11	-.08288	.40917	.72372
V12	.50192	.09441	.10918
V13	.76646	.04425	.00330
V14	.27801	.35222	.37342
V15	.19978	.00280	.57849
V16	.47468	.37570	.14341
V17	.53308	.17701	-.11085
V18	.31893	[illegible]	[illegible]
V19	-.07738	.49056	[illegible]
V20	-.02855	.13144	.51891
V21	.15457	.54866	.30095

TRANSFORMATION MATRIX

	FACTOR 1	FACTOR 2	FACTOR 3
FACTOR 1	.74346	.35157	.56891
FACTOR 2	-.44789	.89348	.03312
FACTOR 3	-.49666	-.27943	.82173

Finally the factor-scores corresponding to the rotated factor analysis are obtained by the inner product of $\underset{\sim}{z}$ with the column vectors of $\underset{\sim}{G} + \underset{\sim}{F}\underset{\sim}{P} = \underset{\sim}{S}\ \underset{\sim}{\Lambda}^{-1}\ \underset{\sim}{P}$. These factor score coefficients are given in Table 12.

TABLE 12

FACTOR SCORE COEFFICIENTS
OF ROTATED FACTOR ANALYSIS

	FACTOR 1	FACTOR 2	FACTOR 3
V1	.17793 ✓	-.05499	.03867
V2	.09942	.[illegible]4175 ✓	-.12089
V3	.16539 ✓	.05011	-.08600
V4	-.02541	-.07[illegible]01	.22500 ✓
V5	.15890 ✓	-.09422	.00235
V6	.10384	-.16734	.17442 ✓
V7	.209[illegible] ✓	-.05006	-.01054
V8	.04499	.22204 ✓	-.02707
V9	-.10293	.1[illegible]238	.21454 ✓
V10	-.03937	.24640 ✓	-.00718
V11	-.10481	.09429	.26283 ✓
V12	.13776	-.01027	-.01092
V13	.22679 ✓	.00335	-.08113
V14	.03951	.07102	.07442
V15	-.00370	.03271	.20087 ✓
V16	.12727	.10931	-.01727
V17	.16806 ✓	.05728	-.10760
V18	.04291	-.15408	.18670 ✓
V19	-.02414	.16225 ✓	-.03504
V20	-.07517	.02565	.22948 ✓
V21	.00824	.15973 ✓	.06700

The coefficients greater than .15 are marked in Table 12. This can help to provide proper interpretation to the factor-scores. Scattergrams of the factor-scores of the 50 simulated systems are given in Figures 3 and 4.

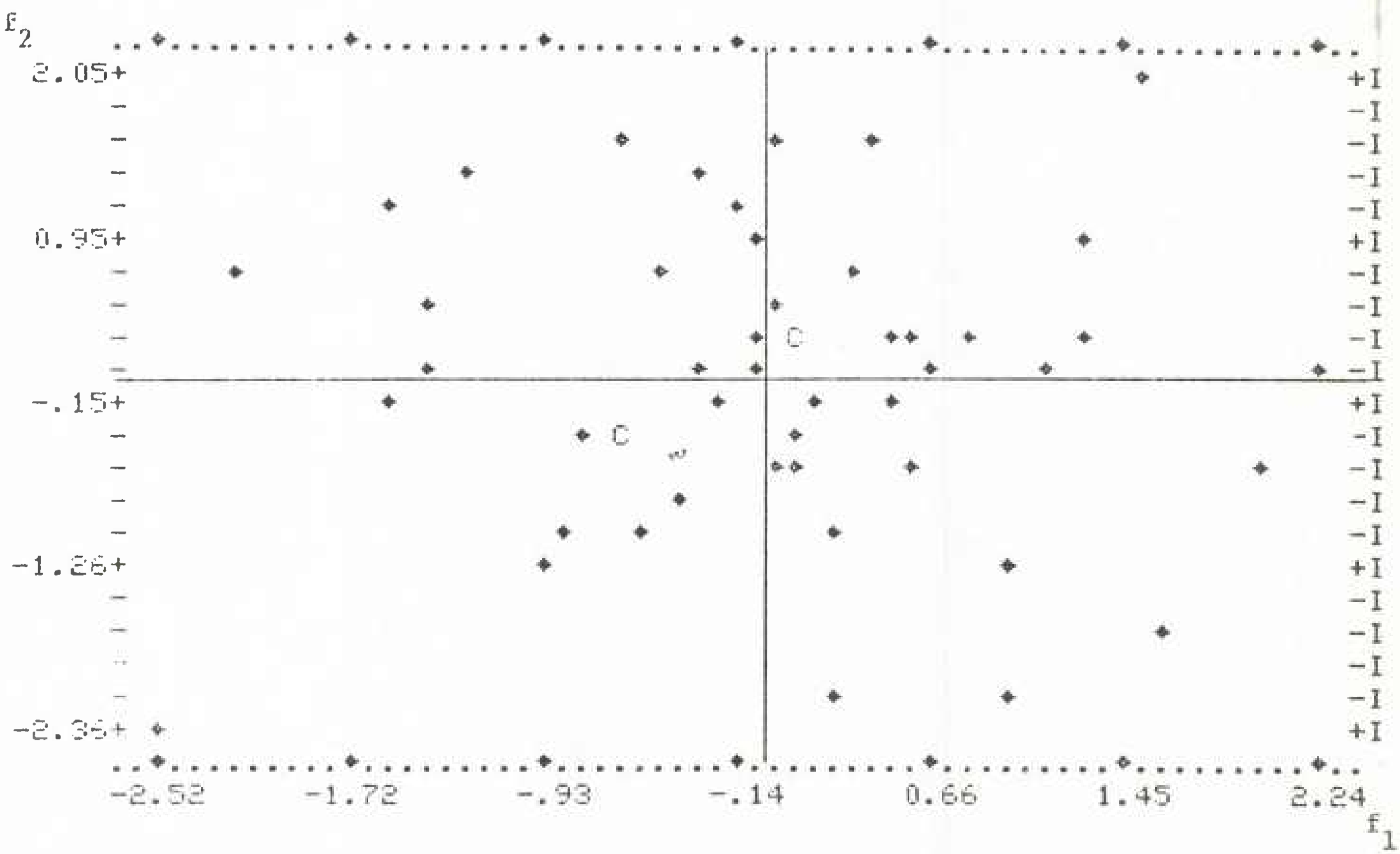


Figure 3. Scattergram of f_2 Versus f_1.

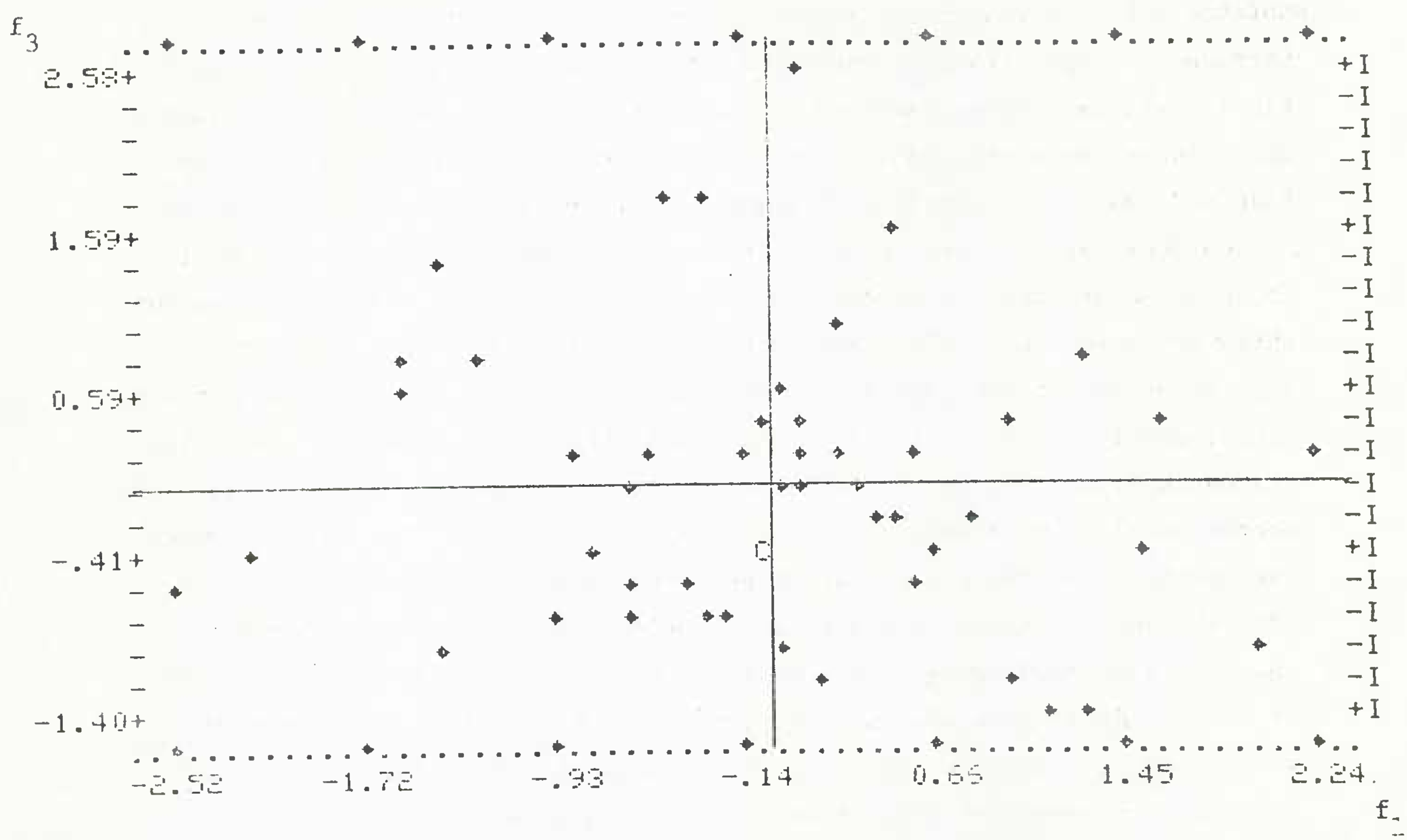


Figure 4. Scattergram of f_3 Versus f_1 .

4. Detecting Deterioration in Readiness, Discrimination and Classification

We have seen in the previous section that the readiness of systems can be represented by principal factors or rotated factors. This is a combined measurement of readiness, which transforms the basic ORI scores and reduces them to a small number of orthogonal factor scores. This representation of the readiness of systems is particularly useful for control purposes. Suppose that we wish to follow the state of readiness of a particular system. We can periodically make observations on the ORI variables and present the corresponding factor scores on the scattergrams given by Figures 3 and 4. Significant deterioration in readiness will be detected by the location of these points in the scattergram. Moreover, if a whole group of points cluster on the scattergram on the negative side of a factor there may be an indication that this group originates from a different population and further analysis (for example, discriminant analysis discussed later) should follow. Such a case was demonstrated in Section 3 and illustrated in Figure 2. In Figure 5 and Figure 6 we illustrate the factor scores obtained by a rotated factor analysis of the 21 ORI variables, when the sample of 50 systems consisted of 25 units from the distribution $N(5\underset{\sim}{1}, 4\underset{\sim}{R})$ and 25 units from the distribution $N(\underset{\sim}{1}, 4\underset{\sim}{R})$. As in Figure 2, the points corresponding to the units in the second subsample are circled. It is seen again that most of the second subsample points are concentrated at the negative part of f_1 . There is a strong indication of a significant difference between the two subsamples. The data consisting of the two subsamples presented in Figures 5 and 6 were further subjected to discriminant analysis of the SPSS. In Appendix III we outline the theory of discriminant analysis. We present here the main ideas and results.

Consider products $\underset{\sim}{\ell}^T \underset{\sim}{v}$ for vectors $\underset{\sim}{v}$ in subsamples 1 and 2. Suppose that the observed vectors in subsample 1 follow the multivariate distribution $N(\underset{\sim}{\mu}_1, 4\underset{\sim}{R})$ and those in subsample 2 are distributed like $N(\underset{\sim}{\mu}_2, 4\underset{\sim}{R})$. The transformed variables $w = \underset{\sim}{\ell}^T \underset{\sim}{v}$ have the normal distributions,

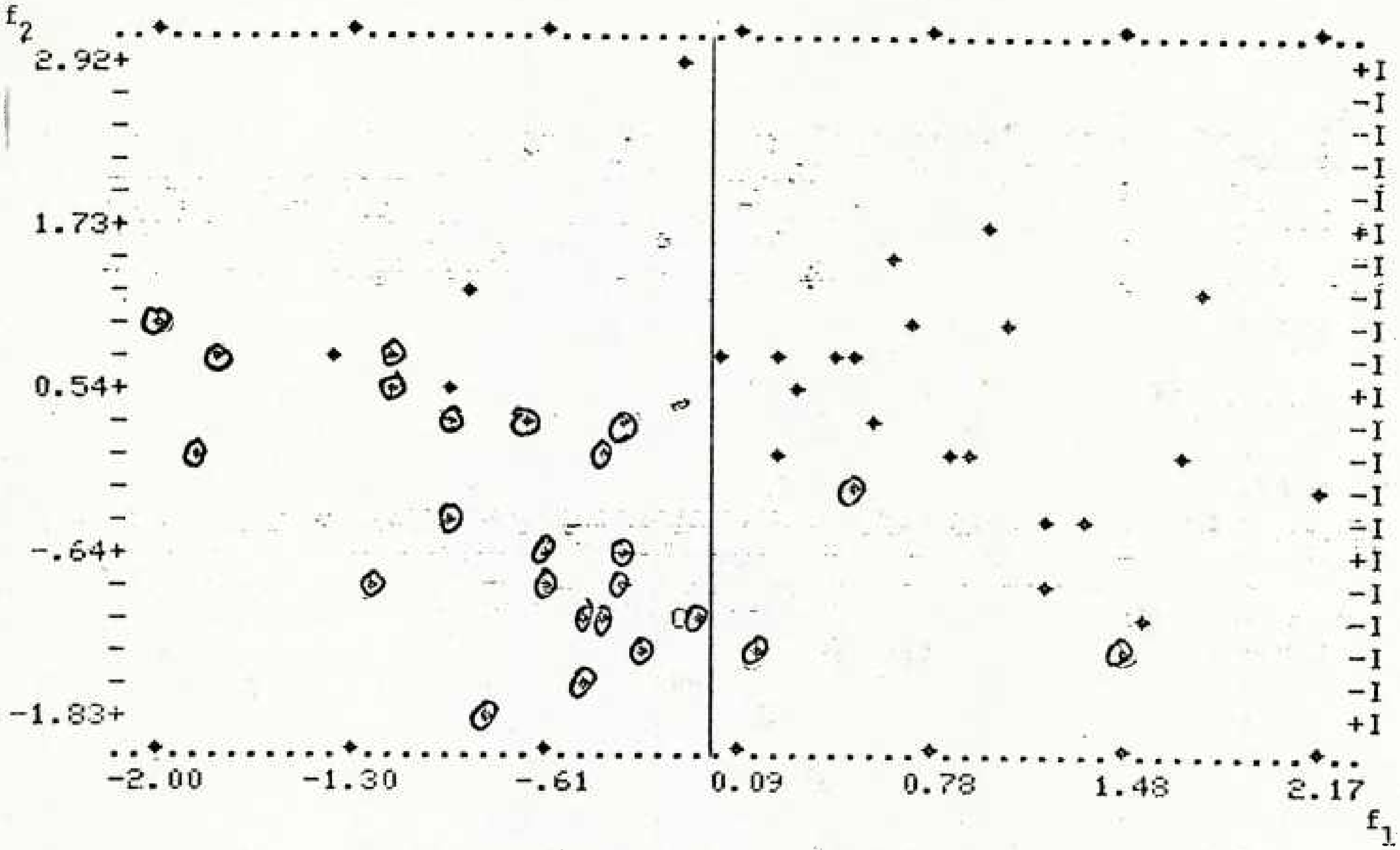


Figure 5. Scattergram of Factor Scores, f_2 Versus f_1, in the Case of Two Subsamples.

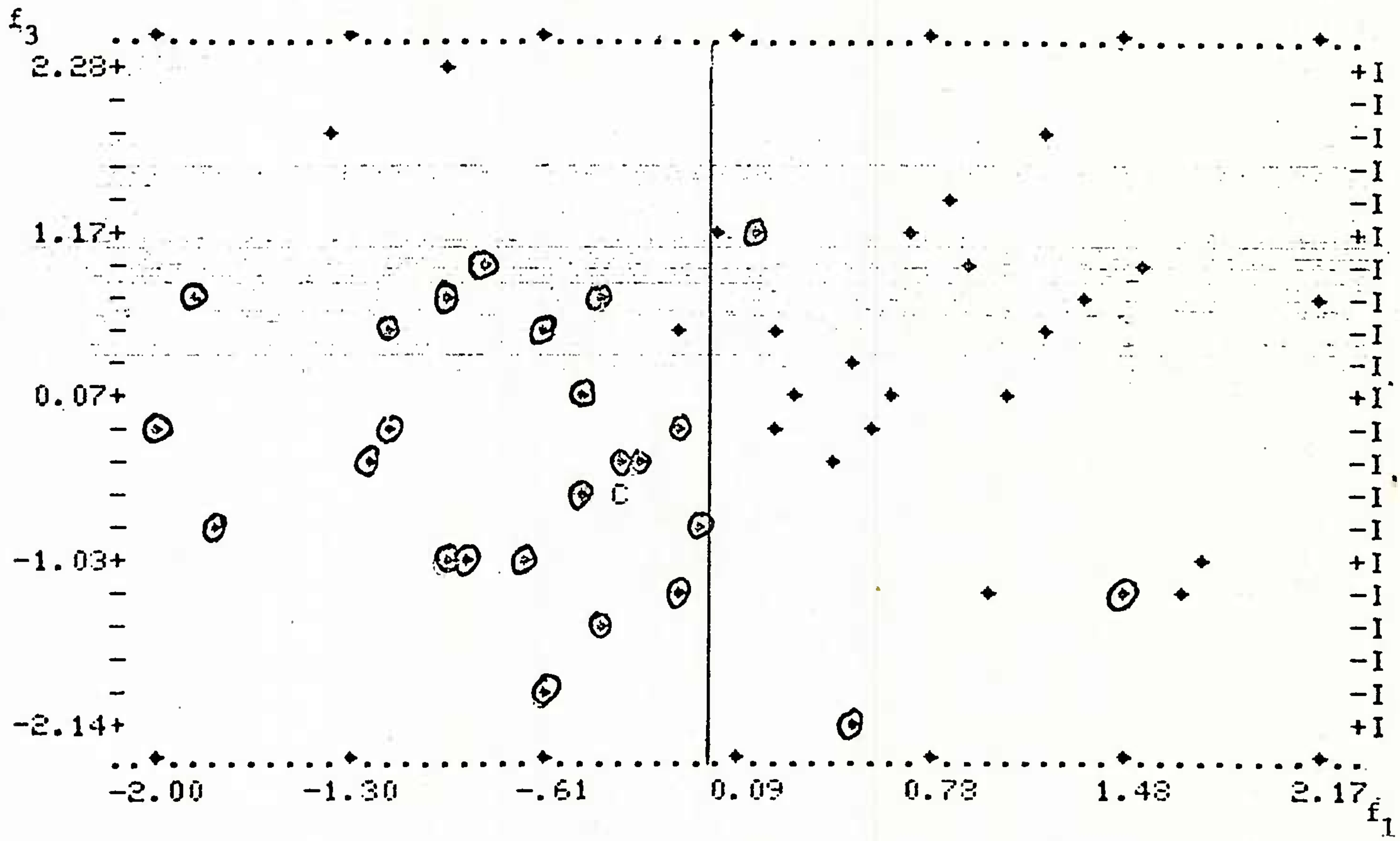


Figure 6. Scattergram of Factor Scores, f_3 Versus f_1, in the Case of Two Subsamples.

$N(\underset{\sim}{\ell}^T \underset{\sim}{\mu}_1, 4\underset{\sim}{\ell}^T \underset{\sim}{R}\underset{\sim}{\ell})$ and $N(\underset{\sim}{\ell}^T \underset{\sim}{\mu}_2, 4\underset{\sim}{\ell}^T R\underset{\sim}{\ell})$, respectively. The problem has thus been reduced to that of discriminating between two univariate samples from normal distributions. The value of $\underset{\sim}{\ell}$ to choose is that which maximizes the F-statistic

$$F = \frac{n_1 n_2}{n_1+n_2} \cdot \frac{(\bar{w}_1-\bar{w}_2)^2}{s_n^2} \tag{4.1}$$

where n_1 and n_2 are the sizes of the subsamples, $\bar{w}_1$ and $\bar{w}_2$ the corresponding sample means of the w variables, and s_n^2 the pooled estimate of the samples within variance of the w variables. The SPSS program provides the optimal vector $\underset{\sim}{\ell}$ and the variables which significantly influence this discrimination (those for which the corresponding components of the optimal are significantly different from zero.) The procedure according to which the variables are chosen to be included in the discriminant analysis is called a "Step-Wise Discriminant Analysis." The reader is referred to Afifi and Azen [1] for a description of the procedure. In addition, the SPSS program performs a classification procedure, which indicates whether or not each one of the vectors in the subsample belongs to the corresponding population (see [1]).

The results of the SPSS discriminant analysis on our simulated data (25 vectors from $N(5\underset{\sim}{1}, 4\underset{\sim}{R})$ and 25 vectors from $N(\underset{\sim}{1}, 4\underset{\sim}{R})$) provide:

- (i) The means of each variable, within the subsamples and total;
- (ii) The standard deviations of each variable, within the subsamples and total;
- (iii) The matrix of within samples intercorrelations (pooled estimates);
- (iv) Optimal coefficients for significant variables;
- (v) Centroids of subsamples, $\bar{w}_1$ and $\bar{w}_2$, for the optimal weights;
- (vi) Table of the classification results.

We provide some of the results in Tables 13 through 18. Table 15 presents the actual classification for every case in the data. For each case we obtain

TABLE 13

SUBSAMPLE MEANS AND STANDARD
DEVIATIONS OF ORI VARIABLES

GROUP COUNTS

	GROUP 1	GROUP 2	TOTAL
COUNT	25.0000	25.0000	50.0000

MEANS

	GROUP 1	GROUP 2	TOTAL
V1	4.7952	.5548	2.6750
V2	4.6792	1.7124	3.1958
V3	4.4428	1.2224	2.8326
V4	5.1508	1.5880	3.3694
V5	4.7988	.8340	2.8164
V6	5.0108	.3352	2.6730
V7	5.1356	.4852	2.8104
V8	4.1724	1.7976	2.9850
V9	3.8664	1.1092	2.4878
V10	4.0592	1.8744	2.9668
V11	3.9188	1.0672	2.4930
V12	4.3948	.2380	2.3164
V13	5.2772	.4372	2.8572
V14	5.3636	1.2324	3.2980
V15	4.4804	.6916	2.5860
V16	4.3080	1.4356	2.8718
V17	4.0388	.7376	2.3882
V18	5.3376	.3844	2.8610
V19	4.3988	.8360	2.6174
V20	4.8772	.9868	2.9320
V21	4.5320	1.4688	3.0004

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	TOTAL
V1	2.3791	1.8818	3.0156
V2	2.3932	2.2611	2.7486
V3	2.1443	2.2132	2.7013
V4	2.2322	1.9654	2.7515
V5	2.3352	1.7899	2.8723
V6	2.5782	2.3847	3.4085
V7	2.5990	1.7232	3.2062
V8	2.2351	2.0562	2.4406
V9	2.0371	1.5795	2.2790
V10	2.3152	1.8633	2.3545
V11	1.9675	1.4987	2.2518
V12	2.3623	1.4143	2.8497
V13	2.3918	1.9273	3.2553
V14	2.5920	2.3591	3.2203
V15	2.2549	1.6892	2.7477
V16	2.6182	1.7687	2.6447
V17	1.7002	1.4128	2.2745
V18	2.4022	2.1118	3.3570
V19	1.6359	1.9947	2.5491
V20	2.5898	1.9335	2.9962
V21	2.5953	2.7217	3.0530

TABLE 14

OPTIMAL DISCRIMINATION WEIGHTS AND SUMMARY
OF CLASSIFICATION RESULTS: TWO-SUBSAMPLES DATA

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V13	-.19723
V15	-.13873
V17	-.23677
V18	-.34179
V19	-.30399

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V13	-.06059
V15	-.05049
V17	-.10410
V18	-.10181
V19	-.11926
CONSTANT	1.15571

CENTROIDS OF GROUPS IN REDUCED SPACE

	FUNC 1
GROUP 1	-.87868
GROUP 2	.87868

DISCRIMINANT ANALYSIS

FILE NONAME
PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP GP. 1	GP. 2
GROUP 1	25.	23.	2.
		92.0%	8.0%
GROUP 2	25.	0.	25.
		.0%	100.0%

PERCENT OF 'GROUPED' CASES CORRECTLY CLASSIFIED: 96.00%

TABLE 15

SPSS CLASSIFICATION ANALYSIS OF TWO-SUBSAMPLES DATA

DISCRIMINANT ANALYSIS

CASE		MISSING	ACTUAL		HIGHEST PROBABILITY			2ND HIGHEST		DISCR
SUBFIL	SEQNUM	VALUES	GROUP		GROUP	P(X/G)	P(G/X)	GROUP	P(G/X)	FUNC t-score
NONA	1.	0	1		1	.997	.993	2	.007	-.60
NONA	2.	0	1		1	.996	.992	2	.008	-.58
NONA	3.	0	1		1	1.000	1.000			-.96
NONA	4.	0	1		1	.941	1.000			-1.39
NONA	5.	0	1		1	1.000	1.000			-1.05
NONA	6.	0	1		1	.772	.755	2	.245	-.13
NONA	7.	0	1		1	.999	1.000			-1.10
NONA	8.	0	1	****	2	.709	.658	1	.342	.08
NONA	9.	0	1		1	1.000	.999	2	.001	-.82
NONA	10.	0	1		1	.994	1.000			-1.18
NONA	11.	0	1		1	.984	.983	2	.017	-.49
NONA	12.	0	1		1	.631	1.000			-1.74
NONA	13.	0	1		1	.912	1.000			-1.45
NONA	14.	0	1		1	.936	.945	2	.055	-.35
NONA	15.	0	1		1	1.000	.997	2	.003	-.70
NONA	16.	0	1		1	1.000	.996	2	.004	-.69
NONA	17.	0	1		1	1.000	.999	2	.001	-.89
NONA	18.	0	1		1	.996	1.000			-1.16
NONA	19.	0	1		1	.657	1.000			-1.72
NONA	20.	0	1		1	.996	1.000			-1.16
NONA	21.	0	1		1	.998	1.000			-1.11
NONA	22.	0	1		1	.999	1.000			-1.03
NONA	23.	0	1		1	.940	1.000			-1.39
NONA	24.	0	1		1	1.000	.998	2	.002	-.75
NONA	25.	0	1	****	2	.991	.988	1	.012	.54

TABLE 15 (Cont'd)

NONA	26.	0	2	2	.996	1.000			1.16
NONA	27.	0	2	2	.983	1.000			1.26
NONA	28.	0	2	2	1.000	1.000			1.00
NONA	29.	0	2	2	.982	1.000			1.27
NONA	30.	0	2	2	1.000	1.000			.96
NONA	31.	0	2	2	.889	1.000			1.46
NONA	32.	0	2	2	1.000	.998	1	.002	.79
NONA	33.	0	2	2	1.000	.999	1	.001	.90
NONA	34.	0	2	2	.962	.966	1	.034	.41
NONA	35.	0	2	2	.826	.828	1	.172	.19
NONA	36.	0	2	2	.997	1.000			1.15
NONA	37.	0	2	2	1.000	.999	1	.001	.81
NONA	38.	0	2	2	1.000	.998	1	.002	.77
NONA	39.	0	2	2	.995	.991	1	.009	.57
NONA	40.	0	2	2	.981	.980	1	.020	.47
NONA	41.	0	2	2	1.000	1.000			.96
NONA	42.	0	2	2	.969	.971	1	.029	.43
NONA	43.	0	2	2	.999	1.000			1.03
NONA	44.	0	2	2	.846	.853	1	.147	.21
NONA	45.	0	2	2	.946	1.000			1.39
NONA	46.	0	2	2	.941	1.000			1.39
NONA	47.	0	2	2	1.000	.998	1	.002	.77
NONA	48.	0	2	2	.884	.896	1	.104	.25
NONA	49.	0	2	2	.999	1.000			1.10
NONA	50.	0	2	2	1.000	1.000			1.0[illegible]

TABLE 16

MEANS AND STANDARD DEVIATIONS IN THREE-SUBSAMPLES DATA

DISCRIMINANT ANALYSIS
FILE NONAME
GROUP COUNTS

	GROUP 1	GROUP 2	GROUP 3	TOTAL
COUNT	20.0000	20.0000	10.0000	50.0000

MEANS

	GROUP 1	GROUP 2	GROUP 3	TOTAL
V1	5.0310	2.7105	.5590	3.2084
V2	4.4715	3.7320	1.5070	3.5828
V3	4.5505	3.3595	1.2470	3.4134
V4	5.3840	3.1215	1.3100	3.6642
V5	5.1495	2.7830	.6580	3.3046
V6	5.3990	2.5260	.6180	3.2934
V7	5.1260	3.0180	.6240	3.3824
V8	4.1415	3.4705	1.8430	3.4134
V9	4.0320	3.2315	.6820	3.0418
V10	4.1140	4.1905	.8970	3.5002
V11	4.3950	3.5620	.6110	3.3050
V12	4.2575	2.6695	.4290	2.8566
V13	5.7155	2.9245	1.2770	3.7114
V14	5.3575	3.1925	1.5460	3.7212
V15	4.4015	2.7170	.8450	3.0164
V16	4.5245	2.8995	1.5770	3.2850
V17	4.4395	2.7470	-.0740	2.8598
V18	4.8730	3.0215	1.5610	3.4700
V19	4.8720	2.5500	.7880	3.1264
V20	4.9720	2.9315	.4300	3.2474
V21	4.1960	3.1010	.7660	3.0720

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	GROUP 3	TOTAL
V1	2.3541	1.7500	1.8069	2.6126
V2	2.5544	1.8466	2.9439	2.5797
V3	2.1972	1.8575	2.4537	2.4088
V4	2.0090	1.9296	2.1671	2.5159
V5	2.6593	1.6224	2.2304	2.7579
V6	2.4060	2.0832	3.0695	3.0280
V7	2.4964	1.6689	1.9315	2.6513
V8	1.8132	2.3947	1.9111	2.2114
V9	2.1418	1.8432	1.5271	2.2529
V10	2.3111	1.9372	1.9083	2.4355
V11	1.9606	1.3739	1.1845	2.1150
V12	2.1020	1.3309	1.6635	2.2195
V13	2.1026	2.0098	2.6209	2.7647
V14	2.1783	2.1930	2.8563	2.7109
V15	1.9903	1.5579	1.6571	2.1825
V16	2.3402	1.7939	1.3197	2.2303
V17	1.7879	1.3360	1.7604	2.2981
V18	1.4830	2.2807	2.4704	2.3706
V19	1.8112	1.4623	1.9360	2.2989
V20	1.7233	1.8880	2.0608	2.4877
V21	2.2607	2.3748	1.9764	2.5470

TABLE 17

SUMMARY OF DISCRIMINANT ANALYSIS STATISTICS THREE-SUBSAMPLES DATA

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
V8	.18405	-.14479
V9	-.12608	-.59925
V10	-.14394	-.78555
V12	-.24636	-.17434
V13	-.04710	.82144
V14	-.13662	.17005
V17	-.41260	-.35913
V18	-.17682	.27193
V19	-.26187	.53829

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
V8	.08323	-.06547
V9	-.05596	-.26599
V10	-.05910	-.32254
V12	-.11100	-.07855
V13	-.01704	.29712
V14	-.05040	.06273
V17	-.17954	-.15628
V18	-.07459	.11471
V19	-.11391	.23415
CONSTANT	1.78926	.36658

CENTROIDS OF GROUPS IN REDUCED SPACE

	FUNC 1	FUNC 2
GROUP 1	-.88931	.40058
GROUP 2	.13352	-.69790
GROUP 3	1.51157	.59464

DISCRIMINANT ANALYSIS
FILE NONAME
PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP GP. 1	GP. 2	GP. 3
GROUP 1	20.	19.	1.	0.
		95.0%	5.0%	.0%
GROUP 2	20.	1.	19.	0.
		5.0%	95.0%	.0%
GROUP 3	10.	0.	1.	9.
		.0%	10.0%	90.0%

PERCENT OF 'GROUPED' CASES CORRECTLY CLASSIFIED: 94.00%

TABLE 18

CLASSIFICATION ANALYSIS OF THREE-SUBSAMPLES DATA

SEQNUM	ACTUAL GROUP		HIGHEST PROBABILITY GROUP	P(X/G)	P(G/X)	2ND HIGHEST GROUP	P(G/X)	DISCRIMINANT SCORES FUNC 1	FUNC 2
1.	1		1	.999	.735	2	.265	-.634	-.283
2.	1		1	.962	.985	2	.015	-.610	1.764
3.	1		1	.997	.987	2	.013	-.769	1.399
4.	1		1	.982	.998	2	.002	-1.119	1.632
5.	1		1	.999	.998	2	.002	-1.386	.599
6.	1		1	.982	.638	2	.360	-.196	.763
7.	1		1	1.000	.948	2	.052	-.779	.466
8.	1	****	2	.999	.910	3	.063	.397	-.003
9.	1		1	1.000	.766	2	.234	-.612	-.113
10.	1		1	.999	.964	2	.036	-1.123	-.331
11.	1		1	.999	.682	2	.318	-.603	-.352
12.	1		1	.948	1.000			-1.704	.800
13.	1		1	.962	1.000			-1.629	.980
14.	1		1	1.000	.891	2	.109	-.620	.441
15.	1		1	1.000	.949	2	.051	-.632	.922
16.	1		1	.997	.819	2	.181	-.367	.837
17.	1		1	1.000	.963	2	.037	-1.109	-.314
18.	1		1	.699	.901	2	.099	-1.304	-1.566
19.	1		1	.534	.998	2	.002	-1.959	-.945
20.	1		1	.998	.996	2	.004	-1.027	1.316
21.	2		2	1.000	.957	1	.031	.245	-.410
22.	2		2	1.000	.873	1	.124	-.031	-.322
23	2		2	.996	.602	1	.397	-.181	.186
24.	2		2	1.000	.965	1	.018	.334	-.509
25.	2		2	1.000	.849	1	.143	.070	.087
26.	2		2	1.000	.984	1	.013	.231	-1.046
27.	2		2	1.000	.972	1	.026	.124	-.908
28.	2		2	.969	.989	1	.011	-.076	-2.055
29.	2		2	.992	.966	3	.033	.676	-1.348
30.	2		2	.995	.832	3	.146	.488	.178
31.	2		2	.940	.913	3	.087	.876	-1.500
32.	2		2	1.000	.956	1	.038	.177	-.496
33.	2		2	1.000	.919	1	.055	.256	.001
34.	2		2	1.000	.947	1	.053	-.040	-.943
35.	2		2	1.000	.754	1	.246	-.300	-.618
36.	2		2	.997	.967	3	.032	.632	-1.216
37.	2		2	1.000	.981	1	.013	.268	-.902
38.	2		2	.998	.634	1	.366	-.419	-.618
39.	2		2	.990	.961	1	.039	-.226	-1.707
40.	2	****	1	.999	.693	2	.307	-.436	.188
41.	3		3	1.000	1.000			1.842	.787
42.	3		3	.997	.992	2	.008	1.148	1.377
43.	3		3	1.000	1.000			1.871	.527
44.	3		3	.987	1.000			1.485	1.839
45.	3		3	.976	1.000			2.175	-.043
46.	3		3	.999	1.000			2.012	.430
47.	3		3	.993	.995	2	.005	1.176	1.559
48.	3		3	.984	.621	2	.378	.893	-.031
49.	3		3	1.000	.994	2	.006	1.244	1.233
50.	3	****	2	.573	.559	3	.441	1.269	-1.731

the likelihood of the observation, given the group $P[\underset{\sim}{X}|G]$ and the Bayes posterior probability of the group, given $\underset{\sim}{X}$. We see in Table 15 two cases in group 1 that behave like the cases of group 2.

The discriminant analysis becomes more complicated, but at the same time more interesting, when the number of subsamples (groups) is greater than two. Generally, if the number of subsamples is k, the analysis is done by computing $k-1$ discriminant functions, which maximize F-statistics that can be obtained by $k-1$ orthogonal vectors $\underset{\sim}{\ell}^{(i)}$ $(i=1,\ldots,k-1)$. The results are then plotted on a 2-dimensional graph, with orthogonal axes representing the first two discriminant functions. We illustrate this in the following example, analyzed by the SPSS program, on three subsamples of sizes $n_1 = 20$, $n_2 = 20$ and $n_3 = 10$. The cases in these subsamples were simulated according to the multinormal distributions $N(5\underset{\sim}{1},4\underset{\sim}{R})$, $N(2\underset{\sim}{1},4\underset{\sim}{R})$ and $N(\underset{\sim}{1},4\underset{\sim}{R})$, respectively. The scattergram in Figure 7 shows that one can well discriminate between these subsamples by means of this analysis. We also learn from the analysis that only variables v_8, v_9, v_{10}, v_{12}, v_{13}, v_{14}, v_{17}, v_{18} and v_{19} contribute significantly to the discrimination.

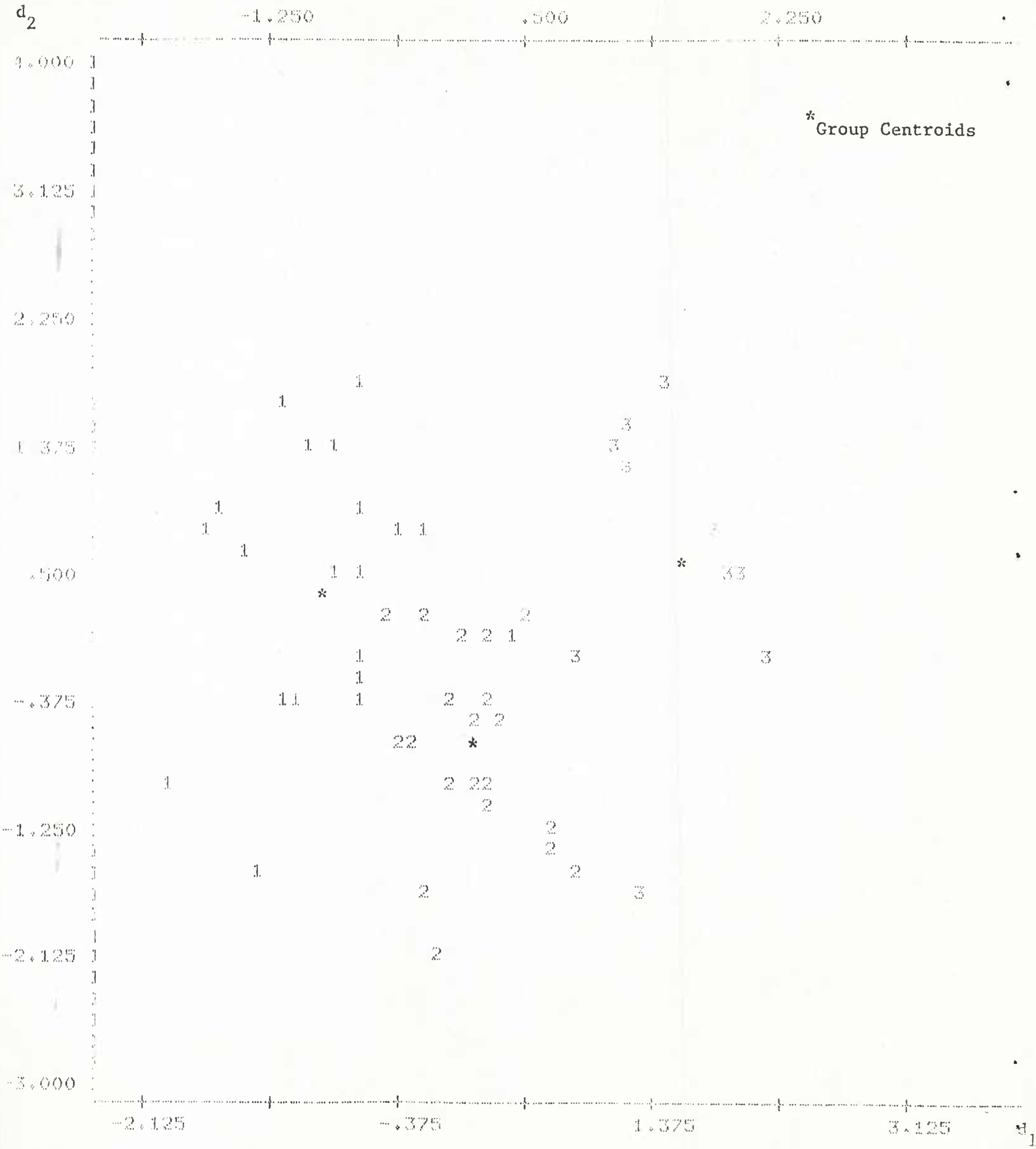


Figure 7. Scattergram of Discriminant Scores.

APPENDIX I

SIMULATING MULTIVARIATE NORMAL VECTORS

The simulation of p-dimensional multivariate normal vectors is based on the following well-known result (see T. W. Anderson [2]).

Let $\mathbf{x} = [x_1,\ldots,x_p]^T$ be distributed like $N(\boldsymbol{\xi},\boldsymbol{\Sigma})$. Suppose that $\mathbf{x}^{(1)} = [x_1,\ldots,x_r]^T$ and $\mathbf{x}^{(2)} = [x_{r+1},\ldots,x_p]^T$, for $1 \leq r < p$. Consider the partition

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}^{(1)} \\ \boldsymbol{\xi}^{(2)} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]$$

where $\boldsymbol{\xi}^{(1)}$ is rx1-dimensional, $\boldsymbol{\xi}^{(2)}$ is (p-r)x1-dimensional, $\boldsymbol{\Sigma}_{11}$ is r x r and $\boldsymbol{\Sigma}_{22}$ is (p-r)x(p-r) . Then, the conditional distribution of $\mathbf{x}^{(2)}$, given $\mathbf{x}^{(1)}$, is the multivariate normal $N(\boldsymbol{\Xi}_{2.1},\boldsymbol{\Sigma}_{22.1})$, where

$$\begin{aligned} \boldsymbol{\xi}_{2.1} &= \boldsymbol{\xi}_2^{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}^{(1)} - \boldsymbol{\xi}^{(1)} \\ \boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12} \end{aligned} \tag{A.1.1}$$

Let $\mathbf{D}$ be a diagonal matrix consisting of the diagonal elements of $\boldsymbol{\Sigma}$, i.e., $\mathbf{D} = \mathrm{diag}(\alpha_{11},\ldots,\alpha_{pp})$. The correlation matrix $\mathbf{R}$ is obtained from $\boldsymbol{\Sigma}$ by the relationship

$$\underset{\sim}{\Sigma} = \underset{\sim}{D}^{1/2}\,\underset{\sim}{R}\,\underset{\sim}{D}^{1/2} \tag{A.1.2}$$

Thus, if $\underset{\sim}{u}$ is a p-dimensional <u>standard</u> multivariate normal vector, i.e., $\underset{\sim}{u} \sim N(\underset{\sim}{0},\underset{\sim}{R})$, then

$$\underset{\sim}{x} = \underset{\sim}{\xi} + \underset{\sim}{D}^{1/2}\underset{\sim}{u} \tag{A.1.3}$$

We consider now the problem of simulating $\underset{\sim}{u}$. Suppose we have a procedure for simulating independent standard normal random variables, $z_1, z_2, \ldots$; $z_i \sim N(0,1)$, $i = 1,2,\ldots$ we set $u_1 = z_1$. The conditional distribution of u_2 , given u_1 is $N(\rho_{12}u_1, 1-\rho_{12}^2)$, where ρ_{12} is the correlation between u_1 and u_2 . Accordingly, given u_1 , we obtain u_2 by the formula

$$u_2 = \rho_{12}u_1 + \left(1-\rho_{12}^2\right)^{1/2} z_2 \tag{A.1.4}$$

Suppose we have already simulated the values of $u_1, u_2, \ldots, u_k$ $(k=1,2,\ldots,p-1)$. Let $\underset{\sim}{u}^{(k)}$ designate this vector. The conditional distribution of u_{k+1} , given $\underset{\sim}{u}^{(k)}$ is obtained according to (A.1.1) in the following manner. Let $\underset{\sim}{R}_k$ be the correlation matrix of $\underset{\sim}{u}^{(k)}$. Consider the partition

$$\underset{\sim}{R}^{(k+1)} = \left[\begin{array}{c|c} \underset{\sim}{R}^{(k)} & \underset{\sim}{\rho}^{(k)} \\ \hline \underset{\sim}{\rho}^{(k)T} & 1 \end{array}\right] \tag{A.1.5}$$

where $\underset{\sim}{\rho}^{(k)}$ is the vector of correlations between u_{k+1} and the components of $\underset{\sim}{u}^{(k)}$. Hence, the conditional distribution of u_{k+1} , given $\underset{\sim}{u}^{(k)}$ is normal with conditional expectation

$$\eta_{k+1} = E\left\{u_{k+1} \mid \underset{\sim}{u}_k\right\} = (\underset{\sim}{\rho}^{(k)})^T (\underset{\sim}{R}^{(k)})^{-1} \underset{\sim}{u}^{(k)} \tag{A.1.6}$$

and conditional variance

$$\tau_{k+1}^2 = 1 - (\underset{\sim}{\rho}^{(k)})^T(\underset{\sim}{R}^{(k)})^{-1}(\underset{\sim}{\rho}^{(k)}) \qquad \text{(A.1.7)}$$

Finally, u_{k+1} is determined by the formula

$$u_{k+1} = \eta_{k+1} + \tau_{k+1} \cdot z_{k+1} \,, \quad k = 1,2,\ldots,p-1 \qquad \text{(A.1.8)}$$

An alternative approach for simulating the vector $\underset{\sim}{u}$ is as follows: Let $\underset{\sim}{C}$ be a non-singular matrix such that $\underset{\sim}{R} = \underset{\sim}{C}\underset{\sim}{C}^T$. Simulate p i.i.d. $N(0,1)$ random variables $z_1,\ldots,z_p$. Let $\underset{\sim}{z} = [z_1,\ldots,z_p]^T$. Then

$$\underset{\sim}{u} = \underset{\sim}{C}\,\underset{\sim}{z} \qquad \text{(A.1.9)}$$

has the standard multivariate normal distribution $N(\underset{\sim}{0},\underset{\sim}{R})$. The matrix $\underset{\sim}{C} = \underset{\sim}{B}\underset{\sim}{\Lambda}^{1/2}$, where $\underset{\sim}{B}$ is a matrix whose columns are the p eigenvectors of $\underset{\sim}{R}$ and $\underset{\sim}{\Lambda}$ is a diagonal matrix of the eigenvalues of $\underset{\sim}{R}$.

APPENDIX II

PRINCIPAL AND ROTATED FACTOR ANALYSIS

Let $\underset{\sim}{\mu} \sim N(\underset{\sim}{0},\underset{\sim}{R})$ be a standard multivariate normal vector. The distribution of $\underset{\sim}{\ell}^T\underset{\sim}{\mu}$ is like that of $N(0,\underset{\sim}{\ell}^T\underset{\sim}{R}\underset{\sim}{\ell})$. Let $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p > 0$ be the eigenvalues of $\underset{\sim}{R}$. We wish to determine a vector (functional) $\underset{\sim}{\ell}$, with length $||\ell|| = 1$, which maximizes the variance of $\underset{\sim}{\ell}^T\underset{\sim}{\mu}$. The Lagrangian is

$$f(\underset{\sim}{\ell},\lambda) = \underset{\sim}{\ell}^T\underset{\sim}{R}\underset{\sim}{\ell} - \lambda(\underset{\sim}{\ell}^T\underset{\sim}{\ell}-1) \tag{A.2.1}$$

Differentiating $f(\underset{\sim}{\ell},\lambda)$ with respect to $\underset{\sim}{\ell}$ yields the eigenstructure equation

$$\underset{\sim}{R}\underset{\sim}{\ell}^o = \lambda_1\underset{\sim}{\ell}^o \tag{A.2.2}$$

Notice that

$$\underset{\sim}{\ell}^T\underset{\sim}{R}\underset{\sim}{\ell} = \lambda\underset{\sim}{\ell}^T\underset{\sim}{\ell} = \lambda$$

Thus, $\underset{\sim}{\ell}^{(1)}$ is an eigenvector of $\underset{\sim}{R}$, of unit length, corresponding to the largest eigenvalue of $\underset{\sim}{R}$, namely to λ_1 which is the variance of $\underset{\sim}{\ell}^{(1)T}\underset{\sim}{\mu}$.

Similarly, let $\underset{\sim}{\ell}^{(2)},\ldots,\underset{\sim}{\ell}^{(p)}$ be the eigenvectors of unit length of $\underset{\sim}{R}$, corresponding to $\lambda_2,\ldots,\lambda_p$. Notice that the variance of $(\underset{\sim}{\ell}^{(i)})^T\underset{\sim}{\mu}$ is λ_i $(i=1,\ldots,p)$ and that

$$\mathrm{cov}(\underset{\sim}{\ell}^{(i)T}\underset{\sim}{\mu},\underset{\sim}{\ell}^{(j)T}\underset{\sim}{\mu}) = 0 \ , \quad \text{all} \quad i \neq j \ . \tag{A.2.2}$$

Indeed, if $\underset{\sim}{R}_i = \lambda_i\underset{\sim}{\ell}^{(i)}\underset{\sim}{\ell}^{(i)T}$, $i = 1,\ldots,p$, then the spectral decomposition of $\underset{\sim}{R}$ is

$$\underset{\sim}{R} = \sum_{j=1}^{p} \tilde{\underset{\sim}{R}}_j \tag{A.2.3}$$

Furthermore, for any $i \neq j$

$$\begin{aligned}\operatorname{cov}(\underset{\sim}{\ell}^{(i)T}\underset{\sim}{u}\ ,\ \underset{\sim}{\ell}^{(j)T}\underset{\sim}{u}) &= \underset{\sim}{\ell}^{(i)T}\underset{\sim}{R}\underset{\sim}{\ell}^{(j)} \\ &= \sum_{k=1}^{p} \underset{\sim}{\ell}^{(i)T}\underset{\sim}{R}_k\underset{\sim}{\ell}^{(j)} = 0\end{aligned} \tag{A.2.4}$$

Let $\underset{\sim}{B} = (\underset{\sim}{\ell}^{(1)},\ldots,\underset{\sim}{\ell}^{(p)})$ be an orthogonal matrix with columns which are the eigenvectors of $\underset{\sim}{R}$. The distribution of

$$\underset{\sim}{f} = \underset{\sim}{\Lambda}^{-1/2}\underset{\sim}{B}^T\underset{\sim}{u} \tag{A.2.5}$$

is like that of $N(\underset{\sim}{0},\underset{\sim}{I})$; where $\underset{\sim}{\Lambda} = \operatorname{diag}(\lambda_1,\ldots,\lambda_p)$. Indeed, $\underset{\sim}{B}^T\underset{\sim}{u} \sim N(\underset{\sim}{0},\underset{\sim}{B}^T\underset{\sim}{R}\underset{\sim}{B})$. But $\underset{\sim}{B}^T\underset{\sim}{R}\underset{\sim}{B} = \underset{\sim}{\Lambda}$. The components of $\underset{\sim}{f}$ are called the <u>principal factor scores</u>, corresponding to $\underset{\sim}{u}$. The orthogonal transformation of $\underset{\sim}{u}$, given by (A.2.5) yields independent standard normal random variables.

Since trace $\underset{\sim}{R}$ = trace $\underset{\sim}{\Lambda} = \sum_{i=1}^{p} \lambda_i = p$, the ratio λ_i/p $(i=1,\ldots,p)$ is the proportion of the total variance of $\underset{\sim}{u}$ accounted for by f_i $(i=1,\ldots,p)$. If we choose only the first m $(1 \leq m < p)$ eigenvectors of $\underset{\sim}{R}$, corresponding to $\lambda_1 \geq \cdots \geq \lambda_m$, and define $\underset{\sim}{B}_{(m)} = [\underset{\sim}{\ell}^{(1)},\ldots,\underset{\sim}{\ell}^{(m)}]$, $\underset{\sim}{\Lambda}_{(m)} = \operatorname{diag}(\lambda_1,\ldots,\lambda_m)$, the transformation $\underset{\sim}{\Lambda}_{(m)}^{-1/2}\underset{\sim}{B}_{(m)}\underset{\sim}{u}$ yields the first m components of $\underset{\sim}{f}$. The concepts of communality and the nature of rotated factor analysis was explained in Section 3.

APPENDIX III

DISCRIMINANT ANALYSIS

Suppose there are k groups (from different populations). We assume that these groups constitute random samples (of systems) of size n_i $(i=1,\ldots,k)$. Let $N = \sum_{i=1}^{k} n_i$ be the total number of observations. Each observation is represented by a p-dimensional random vector, $\underset{\sim}{x}$, having a multivariate normal distribution like $N(\underset{\sim}{\mu}_i,\underset{\sim}{\Sigma})$.

We define the following sample statistics.

The within groups sample covariance matrices

$$\underset{\sim}{S}^{(i)} = [s_{jj'}^{(i)};j,j'=1,\ldots,p] \ , \quad i = 1,\ldots,k \tag{A.3.1}$$

where

$$s_{jj'}^{(i)} = \frac{1}{n_i-1} \sum_{\ell=1}^{n_i} (x_{j\ell}^{(i)}-\bar{x}_j^{(i)})(x_{j'\ell}^{(i)}-\bar{x}_{j'}^{(i)}) \tag{A.3.2}$$

is the sample covariance between x_j and $x_{j'}$ within the ith group $(i=1,\ldots,k;j,j'=1,\ldots,p)$; and $\bar{x}_j^{(i)}$ $(i=1,\ldots,k;j=1,\ldots,p)$ is the ith group mean of the jth variable. The pooled within covariance matrix is

$$\underset{\sim}{W} = \frac{1}{N-k} \sum_{i=1}^{k} (n_i-1)\, \underset{\sim}{S}^{(i)} \tag{A.3.3}$$

Similarly, we define the between samples covariance matrix $\underset{\sim}{B} = [b_{jj'};j,j'=1,\ldots,p]$, where

$$b_{jj'} = \frac{1}{k-1} \sum_{i=1}^{k} n_i(\bar{x}_j^{(i)}-\bar{\bar{x}}_j)(\bar{x}_{j'}^{(i)}-\bar{\bar{x}}_j) \tag{A.3.4}$$

with $\bar{\bar{x}}_j = \frac{1}{N} \sum_{i=1}^{k} n_i\, \bar{x}_j^{(i)}$, as the grand mean of the jth variable $j = 1,\ldots,p$.

If $L = \underset{\sim}{\ell}^T \underset{\sim}{x}$, an F-test of the significance of differences between the centroids of the k groups is

$$F_{\underset{\sim}{\ell}} = \frac{\underset{\sim}{\ell}^T \underset{\sim}{B} \underset{\sim}{\ell}}{\underset{\sim}{\ell}^T \underset{\sim}{W} \underset{\sim}{\ell}} \tag{A.3.5}$$

We determine $\underset{\sim}{\ell}$ to maximize $F_{\underset{\sim}{\ell}}$. Notice that $\underset{\sim}{W}$ is non-singular with probability one and the rank of $\underset{\sim}{B}$ is $k-1$. Differentiating $F_{\underset{\sim}{\ell}}$ with respect to $\underset{\sim}{\ell}$ yields the gradient

$$\nabla_{\underset{\sim}{\ell}} F_{\underset{\sim}{\ell}} = \frac{2}{(\underset{\sim}{\ell}' \underset{\sim}{W} \underset{\sim}{\ell})^2} \left[\left(\underset{\sim}{\ell}^T \underset{\sim}{W} \underset{\sim}{\ell} \right) \underset{\sim}{B} \underset{\sim}{\ell} - \left(\underset{\sim}{\ell}^T \underset{\sim}{B} \underset{\sim}{\ell} \right) \underset{\sim}{W} \underset{\sim}{\ell} \right] \tag{A.3.6}$$

Thus, the vector $\underset{\sim}{\ell}$ which maximizes $F_{\underset{\sim}{\ell}}$ should satisfy the equation

$$B \underset{\sim}{\ell} = F_{\underset{\sim}{\ell}} \underset{\sim}{W} \underset{\sim}{\ell} \tag{A.3.7}$$

or by multiplying (A.3.7) by $\underset{\sim}{W}^{-1}$ we obtain the eigenstructure

$$(\underset{\sim}{W}^{-1} \underset{\sim}{B}) \underset{\sim}{\ell} = F_{\underset{\sim}{\ell}} \underset{\sim}{\ell} \tag{A.3.8}$$

where $\underset{\sim}{W}^{-1} \underset{\sim}{B}$ is positive semi-definite, of rank $k-1$.

Let $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{k-1} > 0$ be the ordered positive eigenvalues of $\underset{\sim}{W}^{-1} \underset{\sim}{B}$. Let $\underset{\sim}{\ell}^{(i)}$, $i = 1, \ldots, k-1$ be the corresponding eigenvectors (of length 1). The functions

$$d_i = \underset{\sim}{\ell}^{(i)T} \underset{\sim}{x} , \qquad i = 1, \ldots, k-1 \tag{A.3.9}$$

are the discriminant scores corresponding to the vector $\underset{\sim}{x}$. The eigenvalue λ_1 is the maximal $F_{\underset{\sim}{\ell}}$ statistic, λ_2 is the second largest, etc. In Section 4, the discriminant vectors (functions) $\underset{\sim}{\ell}^{(i)}$ as well as the discriminant scores are presented. In the case of $k \geqq 3$ it is customary to present the points (d_1, d_2) of each case graphically, on orthogonal systems of axes. Note that although $\underset{\sim}{\ell}^{(i)}$ is orthogonal to $\underset{\sim}{\ell}^{(i')}$, $i, i' = 1, \ldots, k-1$, the random variables d_i are not necessarily uncorrelated.

APPENDIX IV

COMPUTER PROGRAMS

In the present appendix we provide two computer programs: (i) Program SIMU and (ii) Program SNOR2.

Program SIMU, in FORTRAN, provides a complete package of simulating 6-variable multinormal random vectors, according to given distribution means, standard deviations, and correlation matrix. The simulated random vectors are then subjected to principal factor analysis. The program computes and prints the simulated sample means, standard deviations, and correlation matrix; it applies a library subroutine program to determine the eigenvalues and unit length eigenvectors and prints these statistics. After this the program computes the principal factor scores and applies a two-dimensional graphics subroutine to present the scattergram of f_2 versus f_1.

Program SNOR2 was designed to simulate 21-variable multinormal vectors and prepare the result for use in subsequent SPSS analysis by batch process. The computations of all subprograms were performed on a Honeywell GE-430 time-sharing computer. The results were merged in a proper manner and copied into a tape from which the batch was then read into a UNIVAC 1108 computer for the SPSS analysis. In the following we provide specific explanation of the two programs and copies of the programs.

4.1 PROGRAM SIMU

Block	Lines	Designation
1	100-130	Specific merging of library subroutine functions for (i) Solving linear equations: LINEQ (ii) Finding eigenvalues and eigenvectors of a symmetric matrix: EIG1 (iii)-(iv) Simulating standard normal variates: XNORM, RANDX.
2	140-230	Dimension statements and definition of constants.

Block	Lines	Designation
3	250-320	Reading means, standard-deviations and correlations parameters from a data file "CORR".
4	330-390	Setting initial zero values for sums, sum of squares, and sum of products of simulated variables.
5	400-870	Recursive simulation of 50 6-dimensional multinormal vectors and computing their sums, sum of squares, and sum of products.
6	880-970	Computing sample means, sample standard deviations, and sample correlations of the simulated vectors.
7	980-1080	Printing the sample means, standard-deviations, and correlation matrix.
8	1090-1200	Preparation for computation of eigenvalues and eigenvectors of sample correlation matrix.
9	1210	Computation of eigenvalues and eigenvectors of sample correlation matrix.
10	1220-1310	Printing of eigenvalues and eigenvectors.
11	1320-1440	Ordering the eigenvectors corresponding to the M largest eigenvalues.
12	1450-1600	Computing and printing the M factor scores.
13	1610	Graphing the 2-dimensional scattergram of f_2 versus f_1 .
14	1640-end	Subroutine Program PLT1(N,•,•)

4.2 PROGRAM SNOR2

Block	Lines	Designation
1	100-130	Merging library subroutine programs for (i) Solving linear equations: LINEQ (ii)-(iii) Simulating standard normal deviates: XNORM, RANDX.

Block	Lines	Designation
2	140-240	Dimension statements and definition of constants.
3	250-320	Reading means, standard-deviations and correlations of 21 variables, from data file "CORR".
4	330-390	Setting initial zero values for sums, sum of squares, and sum of products.
5	392-875	Recursive simulations of 21-dimensional multinormal vectors and writing the results directly into computer file "RSCORE".
6	880-1085	Computing the sample means, standard deviations and intercorrelations of the 21 variables and writing the results into files "SMNS" and "MATRIX".
7	1090	END.

PROGRAM SIMU

```
SIMU      15:38   CHI    SAT 10/14/78

$LIB,LINEQ,,,***
$LIB,EIG1,,,***
$LIB,XNORM,,,***
$LIB,RANDX,,,***
      DIMENSION X(6),R(6,6),AV(6),SD(6),Z(6),C(5,5),B(5),E(5)
      DIMENSION S(6),P(6,6),EM(6),ESD(6),ER(6,6)
      DIMENSION W(50,6),EG(6,6),TEMP1(6),TEMP2(6),T(6),EH(21)
      DIMENSION EL(4),IT(4),Y(50,3)
      DIMENSION AM(50),F1(50),F2(50)
      DIMENSION EGG(6,3)
      K=6
      M=3
      KK=K-1
      N=50
      AN=N
      U=XNORM(-1.)
      CALL OPENF(1,"CORR")
      READ(1,1)  (AV(I),I=1,K)
    1 FORMAT(6F3.2)
      READ(1,1)  (SD(I),I=1,K)
      DO 3 I=1,K
      READ(1,2)  (R(I,J),J=1,K)
    2 FORMAT(6F3.2)
    3 CONTINUE
      DO 25 I=1,K
      T(I)=0.
      ESD(I)=0.
      DO 30 J=1,K
      P(I,J)=0.
   30 CONTINUE
   25 CONTINUE
      DO 5 I=1,N
  300 U=XNORM(0.)
      Z(1)=U
      X(1)=AV(1)+SD(1)*Z(1)
      W(I,1)=X(1)
      T(1)=T(1)+X(1)
      ESD(1)=ESD(1)+X(1)*X(1)
      U=XNORM(0.)
      Z(2)=R(1,2)*Z(1)+U*SQRT(1.-R(1,2)*R(1,2))
      X(2)=AV(2)+SD(2)*Z(2)
      W(I,2)=X(2)
      T(2)=T(2)+X(2)
      ESD(2)=ESD(2)+X(2)*X(2)
      DO 6 J=3,K
```

```
      JJ=J-1
      DO 7  L1=1,JJ
      DO 8 L2=1,JJ
      C(L1,L2)=R(L1,L2)
    8 CONTINUE
    7 CONTINUE
      DO 9 L1=1,JJ
      B(L1)=R(L1,J)
      E(L1)=B(L1)
    9 CONTINUE
      CALL LINEQ(C,B,JJ,1,KK)
      RSQ=0.
      H=0.
      DO 10 L1=1,JJ
      H=H+B(L1)*Z(L1)
      RSQ=RSQ+B(L1)*E(L1)
   10 CONTINUE
      U=XNORM(0.)
      Z(J)=H+U*SQRT(1.-RSQ)
      X(J)=AV(J)+SD(J)*Z(J)
      W(I,J)=X(J)
      T(J)=T(J)+X(J)
      ESD(J)=ESD(J)+X(J)*X(J)
    6 CONTINUE
      DO 66 L=1,K
      DO 67 J=1,K
     P(L,J)=P(L,J)+X(L)*X(J)
   67 CONTINUE
   66 CONTINUE
     5 CONTINUE
      DO 70 I=1,K
      EM(I)=T(I)/AN
      VI=(AN*ESD(I)-T(I)*T(I))/(AN*(AN-1.))
      ESD(I)=SQRT(VI)
      DO 71 J=1,I
      QIJ=(AN*P(I,J)-T(I)*T(J))/(AN*(AN-1.))
      ER(I,J)=QIJ/(ESD(I)*ESD(J))
      ER(J,I)=ER(I,J)
   71 CONTINUE
   70 CONTINUE
      PRINT 80,
   80 FORMAT(5X,"SAMPLE MEANS AND STAND.DEV.",//)
        PRINT 85,(EM(I),I=1,K)
        PRINT 85,(ESD(I),I=1,K)
    85 FORMAT(5X,6F10.3)
        PRINT 90,
    90 FORMAT(//,5X,"CORRELATIONS MATRIX",//)
        DO 91 I=1,K
```

```
      PRINT 92,(ER(I,J),J=1,K)
   92 FORMAT(5X,6F10.6)
   91 CONTINUE
      DO 93 I=1,K
      DO 94 J=I,K
      IF(I-1) 95,95,96
   95 IJ=J
      GO TO 99
   96 IF(I-2) 97,97,98
   97 IJ=5+J
      GO TO 99
   98 IJ=(I-1)*K-I*(I-1)/2+J
   99 EH(IJ)=ER(I,J)
   94 CONTINUE
   93 CONTINUE
      CALL EIG1(EH,EG,K,2.0E-12,TEMP1,TEMP2,1,K)
      PRINT 100,
  100 FORMAT(//,5X,"EIGENVALUES OF CORREL. MATRIX",//)
      PRINT 110,(EH(J),J=1,K)
  110 FORMAT(5X,6F10.6)
      PRINT 120,
  120 FORMAT(//,"EIGENVECTORS",//)
      DO 130 I=1,K
      PRINT 140,(EG(I,J),J=1,K)
  140 FORMAT(5X,6F10.6)
  130 CONTINUE
     EL(1)=EH(1)
      IT(1)=1
      DO 160 L=1,M
      DO 150 I=1,K
      IF(EL(L)-EH(I)) 145,145,150
  145 EL(L)=EH(I)
      IT(L)=I
  150 CONTINUE
      LTI=IT(L)
      DO 165 J=1,K
      EGG(J,L)=EG(J,LTI)
  165 CONTINUE
      EH(LTI)=0.
     IT(L+1)=1
      EL(L+1)=EH(1)
  160 CONTINUE
      PRINT 200,
  200 FORMAT(//,5X,"FACTOR SCORES",//)
      DO 170 I=1,M
      AM(I)=I
      DO 175 L=1,M
      Y(I,L)=0.
```

```
      DO 180 J=1,K
      Y(I,L)=Y(I,L)+EGG(J,L)*(W(I,J)-EM(J))/ESD(J)
  180 CONTINUE
      Y(I,L)=Y(I,L)/SQRT(EL(L))
  175 CONTINUE
      PRINT 210,I,(Y(I,L),L=1,M)
 210 FORMAT(5X,I4,3F10.4)
      F1(I)=Y(I,1)
      F2(I)=Y(I,2)
  170 CONTINUE
      CALL PLT1(N,F1,F2)
      END
      SUBROUTINE PLT1(N,X,Y)
      DIMENSION X(1),Y(1)
      CALL PSET(XMIN,XMAX,YMIN,YMAX)
      CALL SPLT1(N,X,Y,XMIN,XMAX,YMIN,YMAX)
      RETURN
      END
      SUBROUTINE PSET(XMIN,XMAX,YMIN,YMAX)
      XMIN=2.
      XMAX=1.
      YMIN=2.
      YMAX=1.
      RETURN
      END
      SUBROUTINE SPLT1(N,X,Y,XMIN,XMAX,YMIN,YMAX)
      DIMENSION X(1),Y(1),XSCL(61),YSCL(21),ILINE(61)
      IF(XMIN-XMAX) 2,3,3
    3 CALL AMAX(N,X,XMIN,XMAX)
    2 IF(YMIN-YMAX) 7,6,6
    6 CALL AMAX(N,Y,YMIN,YMAX)
    7 CALL PSCL(1,XMIN,SCALX,YMAX,R,XSCL,YSCL,YMIN,XMAX)
      CALL GZRO(0,ILINE)
      DO 12 K=1,21
      CALL GZRO(K,ILINE)
      CALL SKPT(N,K,X,Y,YSCL,XMIN,SCALX,ILINE,1,R)
      CALL TYPIT(K,YSCL,ILINE)
   12 CONTINUE
      CALL GZRO(0,ILINE)
      CALL PSCL(2,XMIN,SCALX,YMAX,R,XSCL,YSCL,YMIN,XMAX)
      RETURN
      END
      SUBROUTINE AMAX(N,X,XMIN,XMAX)
      DIMENSION X(1)
      XMIN=X(1)
      XMAX=X(1)
```

```
      DO 1 J=2,N
      IF(X(J).LT.XMIN) XMIN=X(J)
      IF(X(J).GT.XMAX) XMAX=X(J)
    1 CONTINUE
      RETURN
      END
      SUBROUTINE PSCL(J,XMIN,SCALX,YMAX,R,XSCL,YSCL,YMIN,XMAX)
      DIMENSION XSCL(61),YSCL(21)
      GO TO (5,6),J
    5 SCALX=(XMAX-XMIN)/60.
      SCALY=(YMAX-YMIN)/20.
      R=.5*SCALY
      DO 10 I=1,61
      AI=I-1
      XSCL(I)=XMIN+AI*SCALX
      IF(I-21) 11,11,10
   11 YSCL(I)=YMIN+(20.-AI)*SCALY
   10 CONTINUE
      PRINT 13,
   13 FORMAT(///)
      RETURN
    6 PRINT 50,(XSCL(I),I=1,61,10)
   50 FORMAT(F11.2,6F10.2)
      RETURN
      END
      SUBROUTINE GZRO(K,ILINE)
      DIMENSION ILINE(61),ISYM(3)
      DATA ISYM(1),ISYM(2),ISYM(3)/" ",".","*"/
      L=2
      IF(K) 1,2,1
    1 L=1
    2 DO 3 I=1,61
      ILINE(I)=ISYM(L)
    3 CONTINUE
      GO TO (5,8),L
    8 DO 6 I=1,61,10
    6 ILINE(I)=ISYM(3)
      PRINT 11,ILINE
   11 FORMAT(6X,"..",61A1,"..")
    5 RETURN
      END
      SUBROUTINE SKPT(N,K,X,Y,YSCL,XMIN,SCALX,ILINE,L,R)
      DIMENSION X(1),Y(1),ILINE(61),YSCL(21),ISYM(6)
      DATA IBL,(ISYM(I),I=1,6)/" ","*","2","3","4","5","C"/
      DO 5 I=1,N
      IF(Y(I)-YSCL(K)-R) 2,2,5
    2 IF(Y(I)-YSCL(K)+R) 5,5,4
    4 M=(X(I)-XMIN)/SCALX+1.5
      IF(M) 5,5,7
```

```
    7 IF(M-61) 6,6,5
    6 IF(ILINE(M)-IBL) 8,9,8
    9 ILINE(M)=ISYM(L)
      GO TO 5
    8 ILINE(M)=ISYM(6)
    5 CONTINUE
      RETURN
      END
      SUBROUTINE TYPIT(K,YSCL,ILINE)
      DIMENSION YSCL(21),ILINE(61)
      JS=K+4
      IF(JS-JS/5*5) 60,61,60
   61 PRINT 3,YSCL(K),ILINE
    3 FORMAT(F6.2,"+",1X,61A1,1X,"+I")
      GO TO 12
   60 PRINT 16,ILINE
   16 FORMAT(6X,"-",1X,61A1,1X,"-I")
   12 RETURN
      END

OK
```

PROGRAM SNOR2

```
OLD:SNOR2

OK
LIST

SNOR2      15:44    CHI    SAT 10/14/78

100$LIB,LINEQ,,,***
120$LIB,XNORM,,,***
130$LIB,RANDX,,,***
140        DIMENSION X(21),R(21,21),AV(21),SD(21),Z(21),C(20,20),B(20),E(20)
150        DIMENSION S(21),P(21,21),EM(21),ESD(21),ER(21,21)
160        DIMENSION W(50,21),T(21)
190        K=21
210        KK=K-1
220        N=50
230        AN=N
240        U=XNORM(-1.)
250        CALL OPENF(1,"CORR")
260        READ(1,1) (AV(I),I=1,K)
270      1 FORMAT(21F3.2)
280        READ(1,1) (SD(I),I=1,K)
290        DO 3 I=1,K
300        READ(1,2) (R(I,J),J=1,K)
310      2 FORMAT(21F3.2)
320      3 CONTINUE
330        DO 25 I=1,K
340        T(I)=0.
350        ESD(I)=0.
360        DO 30 J=1,K
370        P(I,J)=0.
380     30 CONTINUE
390     25 CONTINUE
392        DO 1000 I=1,50
394        U=XNORM(0.)
396   1000 CONTINUE
400        DO 5 I=1,N
410        IF(I-21) 300,400,405
420    300 LF=1
425        GO TO 305
430    400 LF=2
435        DO 500 J=1,K
440        AV(J)=AV(J)-2.
```

```
  500 CONTINUE
      GO TO 305
  405 IF(I-41) 305,505,305
  505 LF=3
      DO 605 J=1,K
      AV(J)=AV(J)-2.
  605 CONTINUE
  305 U=XNORM(0.)
      Z(1)=U
      X(1)=AV(1)+SD(1)*Z(1)
      W(I,1)=X(1)
      T(1)=T(1)+X(1)
      ESD(1)=ESD(1)+X(1)*X(1)
      U=XNORM(0.)
      Z(2)=R(1,2)*Z(1)+U*SQRT(1.-R(1,2)*R(1,2))
      X(2)=AV(2)+SD(2)*Z(2)
      W(I,2)=X(2)
      T(2)=T(2)+X(2)
      ESD(2)=ESD(2)+X(2)*X(2)
      DO 6 J=3,K
      JJ=J-1
      DO 7  L1=1,JJ
      DO 8 L2=1,JJ
      C(L1,L2)=R(L1,L2)
    8 CONTINUE
    7 CONTINUE
      DO 9 L1=1,JJ
      B(L1)=R(L1,J)
      E(L1)=B(L1)
    9 CONTINUE
      CALL LINEQ(C,B,JJ,1,KK)
      RSQ=0.
      H=0.
      DO 10 L1=1,JJ
      H=H+B(L1)*Z(L1)
      RSQ=RSQ+B(L1)*E(L1)
   10 CONTINUE
      U=XNORM(0.)
      Z(J)=H+U*SQRT(1.-RSQ)
      X(J)=AV(J)+SD(J)*Z(J)
      W(I,J)=X(J)
      T(J)=T(J)+X(J)
      ESD(J)=ESD(J)+X(J)*X(J)
    6 CONTINUE
      DO 66 L=1,K
      DO 67 J=1,K
      P(L,J)=P(L,J)+X(L)*X(J)
```

```
 67 CONTINUE
 66 CONTINUE
    WRITE(2,100) (W(I,J),J=1,K)
100 FORMAT(7F8.2)
    WRITE(2,105) LF
105 FORMAT(I1)
  5 CONTINUE
    CALL CLOSEF(2,"RSCORE")
    DO 70 I=1,K
    EM(I)=T(I)/AN
    VI=(AN*ESD(I)-T(I)*T(I))/(AN*(AN-1.))
    ESD(I)=SQRT(VI)
    DO 71 J=1,I
    QIJ=(AN*P(I,J)-T(I)*T(J))/(AN*(AN-1.))
    ER(I,J)=QIJ/(ESD(I)*ESD(J))
    ER(J,I)=ER(I,J)
 71 CONTINUE
 70 CONTINUE
    PRINT 80,
 80 FORMAT(5X,"SAMPLE MEANS AND STAND.DEV.",//)
    WRITE(4,85) (EM(I),I=1,K)
    WRITE(4,85) (ESD(I),I=1,K)
 85 FORMAT(5X,7F6.3)
    CALL CLOSEF(4,"SMNS")
    PRINT 90,
 90 FORMAT(//,5X,"CORRELATIONS MATRIX",//)
    DO 91 I=1,K
    WRITE (3,92) (ER(I,J),J=1,K)
 92 FORMAT(8F10.7)
 91 CONTINUE
    CALL CLOSEF(3,"MATRIX")
    END

OK
```

REFERENCES

[1] AFIFI, A. A. and S. P. AZEN (1972). Statistical Analysis: A Computer Oriented Approach. Academic.

[2] ANDERSON, T. W. (1958). Introduction to Multivariate Analysis. Wiley.

[3] BARZILY, Z., W. H. MARLOW, and S. ZACKS (1979). Survey of approaches to readiness. Naval Res. Logist. Quart. (to appear)

[4] COOLEY, W. W. and P. R. LOHNES (1971). Multivariate Data Analysis. Wiley.

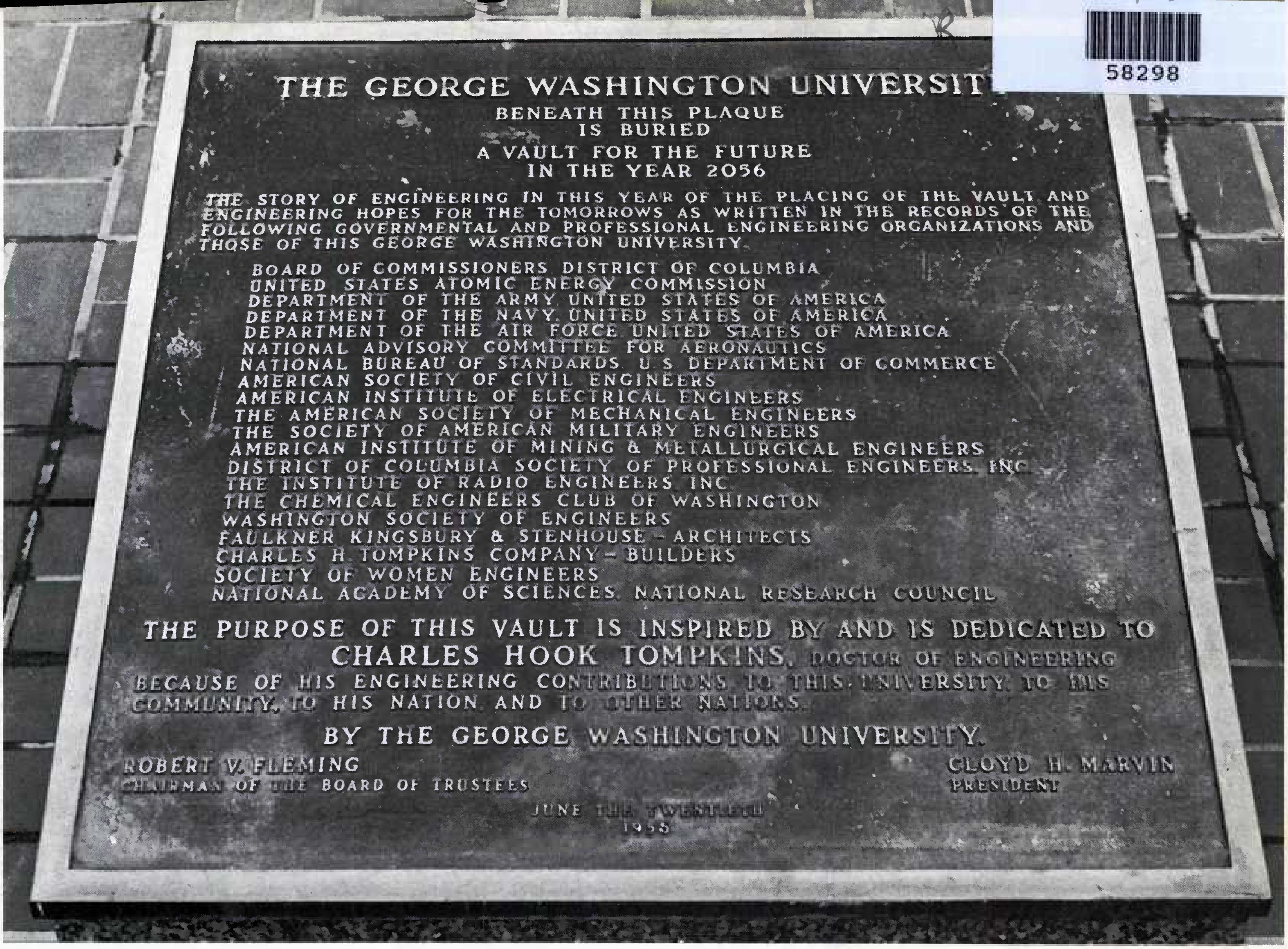
[5] LOCKMAN, R. F., P. H. STOLOFF, B. H. MANHEIMER, J. R. HARDGRAVE and W. F. STORY (1970). Navy Readiness Analysis System Methodology Study. 1 Institute of Naval Studies of the Center for Naval Analyses. Study No. 27. University of Rochester.

[6] OVERALL, J. and C. T. KLETT (1972). Applied Multivariate Analysis. McGraw-Hill.

[7] NORMAN, H. N., C. H. HULL, J. G. JENKINS, K, STEINBRENNER and D. H. BENT (1975). SPSS: Statistical Package for the Social Sciences. 2nd ed. McGraw-Hill.

[8] TATSUOKA, M. M. (1971). Multivariate Analysis: Techniques for Educational and Psychological Research. Wiley.

[9] VAN DE GEER, J. P. (1971). Introduction to Multivariate Analysis for the Social Sciences. Freeman.



To cope with the expanding technology, our society must be assured of a continuing supply of rigorously trained and educated engineers. The School of Engineering and Applied Science is completely committed to this objective.